



The
University
Of
Sheffield.

HIGH DYNAMIC RANGE VISUAL CONTENT COMPRESSION

submitted by

Tian Feng

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

The University of Sheffield
Faculty of Engineering
Department of Electronic and Electrical Engineering

September 21, 2017

ABSTRACT

This thesis addresses the research questions of High Dynamic Range (HDR) visual contents compression. The HDR representations are intended to represent the actual physical value of the light rather than exposed value. The current HDR compression schemes are the extension of legacy Low Dynamic Range (LDR) compressions, by using Tone-Mapping Operators (TMO) to reduce the dynamic range of the HDR contents. However, introducing TMO increases the overall computational complexity, and it causes the temporal artifacts. Furthermore, these compression schemes fail to compress non-salient region differently than the salient region, when Human Visual System (HVS) perceives them differently.

The main contribution of this thesis is to propose a novel Mapping-free visual saliency-guided HDR content compression scheme. Firstly, the relationship of Discrete Wavelet Transform (DWT) lifting steps and TMO are explored. A novel approach to compress HDR image by Joint Photographic Experts Group (JPEG) 2000 codec while backward compatible to LDR is proposed. This approach exploits the reversibility of tone mapping and scalability of DWT. Secondly, the importance of the TMO in the HDR compression is evaluated in this thesis. A mapping-free post HDR image compression based on JPEG and JPEG2000 standard codecs for current HDR image formats is proposed. This approach exploits the structure of HDR formats. It has an equivalent compression performance and the lowest computational complexity compared to the existing HDR lossy compressions (50% lower than the state-of-the-art). Finally, the shortcomings of the current HDR visual

saliency models, and HDR visual saliency-guided compression are explored in this thesis. A spatial saliency model for HDR visual content outperform others by 10% for spatial visual prediction task with 70% lower computational complexity is proposed. Furthermore, the experiment suggested more than 90% temporal saliency is predicted by the proposed spatial model. Moreover, the proposed saliency model can be used to guide the HDR compression by applying different quantization factor according to the intensity of predicted saliency map.

Dedication

Dedicated to my dear grandmother Jingyu Wang, who passed away during my research study. I wish that I can see her one last time. You will always be remembered.

ACKNOWLEDGEMENTS

Firstly, I would thank for the encouragement and supports from my supervisors Dr. Charith Abhayaratne. Also for giving me the opportunity for doing a Ph.D. I could not expect more than have a supervisor who helped so much about my research and a group with high ethical standards.

I am grateful deeply to my colleagues who work in the B28 VIE group with wholehearted advice, close collaboration. I would like to thank Sandipan, Matthew, Jeff, Wattanit, Catherine who were kindly shared their precious study experience and knowledge.

I greatly thank my family, for great help throughout this Ph.D.

Finally, I would like to thank all of my friends for the support and caring.

Contents

| | |
|--|--------------|
| List of Figures | xi |
| List of Tables | xxiii |
| List of Symbols and Acronyms | xxiii |
| 1 Introduction | 1 |
| 1.1 Motivation | 1 |
| 1.2 Research Questions | 3 |
| 1.3 Contributions | 4 |
| 1.4 Thesis Structure | 4 |
| 2 Background and Literature Review | 6 |
| 2.1 Light, Human Visual Adaptation, and Colour Spaces | 7 |
| 2.1.1 Light Quantities and the Contrast | 7 |
| 2.1.1.1 Contrast Ratio | 9 |
| 2.1.2 Human Visual Adaptation | 10 |
| 2.1.3 Color Spaces | 11 |
| 2.1.3.1 Gamma Correction | 13 |
| 2.2 Capturing and Displaying HDR Content | 14 |
| 2.2.1 Capturing of HDR Content | 15 |
| 2.2.1.1 Recovering HDR Radiance maps from Pho- tographs | 15 |
| 2.2.1.2 Native Hardware Solutions | 16 |
| 2.2.2 Display of HDR Content | 17 |

| | | |
|---------|--|----|
| 2.3 | Encoding HDR content | 19 |
| 2.3.1 | HDR Image Lossless Formats | 19 |
| 2.3.1.1 | The Radiance RGBE/ XYZE | 20 |
| 2.3.1.2 | Logluv32 | 21 |
| 2.3.1.3 | EXR | 22 |
| 2.3.1.4 | PFM | 23 |
| 2.3.2 | Encoding HDR image Lossy HDR image Compression . | 24 |
| 2.3.2.1 | JPEG-HDR | 24 |
| 2.3.2.2 | JPEG2000-HDR | 25 |
| 2.3.2.3 | JPEG-XT | 28 |
| 2.3.2.4 | Other HDR Compressions | 28 |
| 2.3.3 | Encoding HDR video | 29 |
| 2.3.3.1 | HDR Video Compression based on AVC (MEPG4/ H.264) | 30 |
| 2.3.3.2 | HDR Video Compression based on HEVC (H.265) | 30 |
| 2.3.4 | Discussion | 31 |
| 2.4 | Tone-Mapping HDR Content | 32 |
| 2.4.1 | Image TMO | 34 |
| 2.4.1.1 | Brightness Reproduction | 34 |
| 2.4.1.2 | Histogram Adjustment | 35 |
| 2.4.1.3 | Photographic Tone Reproduction | 36 |
| 2.4.1.4 | Adaptive Logarithmic Mapping | 39 |
| 2.4.1.5 | Fast Bilateral Filtering | 40 |
| 2.4.1.6 | Computational Complexity Evaluation of TMO | 40 |
| 2.4.2 | Video TMO | 41 |
| 2.4.2.1 | Video TMO artifacts | 41 |
| 2.4.2.2 | Temporally Coherent Local TMO | 43 |
| 2.4.2.3 | Real-Time Noise-Aware TMO | 44 |
| 2.4.3 | Discussion | 44 |
| 2.5 | Visual Quality Evaluation of HDR Content | 45 |

| | | |
|----------|---|-----------|
| 2.5.1 | Objective evaluation of HDR content | 45 |
| 2.5.1.1 | NON-HVS Driven Metrics | 46 |
| 2.5.1.2 | Visible Differences Predictor | 47 |
| 2.5.1.3 | Dynamic Range Independent Image Quality Assessment (DRIQA) | 49 |
| 2.5.1.4 | HDR-VDP2 | 49 |
| 2.5.1.5 | HDR-VQM | 50 |
| 2.5.2 | Subjective Evaluation of HDR Content | 50 |
| 2.5.3 | Discussion | 51 |
| 2.6 | Visual Saliency Models | 52 |
| 2.6.1 | Spatial Saliency Models | 52 |
| 2.6.1.1 | LDR Saliency Models | 52 |
| 2.6.1.2 | HDR Saliency Models | 54 |
| 2.6.2 | Temporal Saliency Models | 57 |
| 2.6.3 | Visual Saliency-Guided Compression | 59 |
| 2.6.4 | Saliency Evaluation Metrics | 59 |
| 2.6.4.1 | Pearson Correlation Coefficient (CC) | 59 |
| 2.6.4.2 | KLD | 60 |
| 2.6.4.3 | ROC | 60 |
| 2.6.5 | Discussion | 60 |
| 2.7 | Evaluation Tools and Data-sets | 61 |
| 2.7.1 | Encoding Tools | 61 |
| 2.7.2 | Image dataset | 61 |
| 2.7.3 | Video Data-set | 63 |
| 2.7.4 | Eye Tracker Fixations | 63 |
| 3 | DWT with TMO built-in HDR Visual Content Compression | 65 |
| 3.1 | Motivation | 65 |
| 3.1.1 | DWT lifting theory | 66 |
| 3.1.1.1 | 1D lifting | 66 |

| | | |
|----------|---|------------|
| 3.1.2 | Extending 1D Haar lifting to 2D Haar lifting | 68 |
| 3.2 | The Proposed method | 70 |
| 3.2.1 | DWT-TMO lifting | 70 |
| 3.2.1.1 | Proposed Haar 2D DWT-TMO lifting | 75 |
| 3.2.1.2 | Extending to 2D $3/5$ and 2D $5/3$ DWT-TMO lifting | 78 |
| 3.3 | Experimental Results | 89 |
| 3.3.1 | HDR test image information | 89 |
| 3.3.2 | Experimental setups | 89 |
| 3.3.2.1 | TMO | 90 |
| 3.3.2.2 | Quantization method | 90 |
| 3.3.2.3 | Entropy Measurement | 91 |
| 3.3.3 | Experimental Results and Discussion | 92 |
| 3.3.4 | Extending Propose DWT-TMO Lifting to HDR Video . | 98 |
| 3.3.4.1 | Tone mapped HDR Video compression artifacts | 100 |
| 3.4 | Summary | 102 |
| 4 | Mapping-Free HDR Visual Content Compression | 104 |
| 4.1 | Motivation | 104 |
| 4.2 | Characteristics Of Existing HDR Image Formats | 105 |
| 4.3 | Proposed mapping-free Post HDR Image Compression Method | 107 |
| 4.3.1 | Computational Complexities | 107 |
| 4.3.2 | Proof of the concept | 108 |
| 4.3.3 | Comparison to Perceptual based Quantization | 111 |
| 4.4 | Proposed JPEG2000-XT HDR Image Compression Method . . | 113 |
| 4.5 | Proposed mapping-free Post HDR Video Compression Method | 113 |
| 4.5.1 | JPEG/ JPEG2000 | 114 |
| 4.5.2 | JPEG2000-XT | 114 |
| 4.5.3 | HEVC | 114 |
| 4.6 | Result and Discussion | 114 |

| | | |
|----------|---|------------|
| 4.6.1 | Experimental Setups | 114 |
| 4.6.2 | Objective Quality Evaluation Metric | 114 |
| 4.6.3 | Experimental Results and Discussion | 115 |
| 4.6.3.1 | Computational Complexities' evaluation | 115 |
| 4.6.3.2 | Image results | 115 |
| 4.6.3.3 | Video results | 121 |
| 4.7 | Summary | 124 |
| 5 | Saliency-Guided HDR Visual Content Compression | 126 |
| 5.1 | Motivation | 126 |
| 5.2 | Saliency Models for HDR Image | 128 |
| 5.2.1 | The Proposed Method | 128 |
| 5.2.1.1 | TMO | 129 |
| 5.2.1.2 | Saliency Detection | 130 |
| 5.2.1.3 | Saliency Fusion | 130 |
| 5.3 | Saliency Model for HDR Video | 132 |
| 5.3.1 | LDR Temporal Models | 132 |
| 5.3.2 | HDR Temporal Models | 134 |
| 5.3.3 | Spatial-Temporal Model Fusion | 138 |
| 5.3.4 | Experimental Results and Discussion | 141 |
| 5.3.4.1 | Computational Complexities evaluation | 142 |
| 5.3.4.2 | Image Results | 143 |
| 5.3.4.3 | Discussion | 147 |
| 5.4 | Saliency Guided HDR Visual Content Encoding | 151 |
| 5.4.1 | Image Encoding | 151 |
| 5.4.2 | Video Encoding | 154 |
| 5.4.3 | Experimental Results and Discussion | 155 |
| 5.4.3.1 | The Costs of Saliency Map | 157 |
| 5.4.3.2 | Image Results | 157 |
| 5.4.3.3 | Video Results | 162 |

| | | |
|----------|-----------------------|------------|
| 5.5 | Summary | 165 |
| 6 | Conclusions | 167 |
| 6.1 | Future work | 170 |

List of Figures

| | | |
|------|---|----|
| 1.1 | (a) A LDR stacks of a given scene in different exposure times and (from shorter to longer exposure times – from top-left to right-bottom) (b) a tone mapped HDR version. | 2 |
| 2.1 | Wavelength image from [41] | 8 |
| 2.2 | CIE standard observer photopic luminous efficiency curve from http://www.cvrl.org | 9 |
| 2.3 | (a) HDR image; (b) TMO result; (c) Gamma corrected result. . | 14 |
| 2.4 | (a) 1/250; (b) 1/60; (c) 1/15; (d) 1; (e) 4 ;(f) 15. secs of exposure times | 17 |
| 2.5 | The inverse camera response curves. | 18 |
| 2.6 | (a) recovered HDR; (b) tone-mapped. | 18 |
| 2.7 | (a) A low light condition tone-mapped HDR frame; (b) zoomed-in version of (a). | 18 |
| 2.8 | From left to right: HDR image reconstructed from only RGB channels of RGBE formats; HDR image reconstructed from RGBE channels of RGBE formats; Tone-mapped HDR image reconstructed from RGBE image; R; G; B; E channels of HDR RGBE image displayed separately | 20 |
| 2.9 | JPEG-HDR coding scheme | 26 |
| 2.10 | (a) HDR image; (b) LDR tone-mapped image; (c) Ratio map which is (a)/(b). | 26 |
| 2.11 | JPEG2000-HDR coding scheme | 27 |

| | | |
|------|---|----|
| 2.12 | JPEG2000-HDR transformed integer image | 27 |
| 2.13 | JPEG-Xt coding scheme | 28 |
| 2.14 | (a) HDR image; (b) TMO result; (c) Gamma corrected result. . | 35 |
| 2.15 | (a) HDR image; (b) TMO result; (c) Gamma corrected result. . | 36 |
| 2.16 | Reinhard global TMO (a) HDR image; (b) TMO result; (c) Gamma corrected result. | 38 |
| 2.17 | Reinhard local TMO(a) HDR image; (b) TMO result; (c) Gamma corrected result. | 38 |
| 2.18 | (a) HDR image; (b) TMO result; (c) Gamma corrected result. . | 39 |
| 2.19 | (a) HDR image; (b) TMO result; (c) Gamma corrected result. . | 41 |
| 2.20 | Example of global flickering artifacts | 42 |
| 2.21 | Example of local flickering artifacts | 42 |
| 2.22 | False colour luminance | 42 |
| 2.23 | Temporal contrast incoherency occurs when illumination changed | 43 |
| 2.24 | False colour luminance | 43 |
| 2.25 | TMO applied | 43 |
| 2.26 | Sweep grating to measure CSF | 48 |
| 2.27 | CSF under different adaptation levels | 48 |
| 2.28 | Architecture of the Itti <i>et al.</i> model | 53 |
| 2.29 | Architecture of the Spatial-Temporal model | 58 |
| 2.30 | (a) Over saturated pixels; (b) zoomed-in tone-mapped version of the example | 64 |
| 3.1 | (a) An one level of DWT decomposition; (b) Lena example of one level of DWT decomposition. | 71 |
| 3.2 | Reinhard Global TMO [108] (a)TMO; (b)ITMO. | 71 |
| 3.3 | (a) HDR image ; (b) Jet colour map of HDR image | 72 |

| | | |
|------|---|-----|
| 3.4 | (a) Tone-mapped HDR image $\mathcal{T}(I)$; (b) Lazy wavelet transformed tone-mapped image $S(\mathcal{T}(I))$; (c) Inverse tone-mapped of (b) $\mathcal{T}'(S(\mathcal{T}(I)))$; (d) Inverse Lazy wavelet transformed of (c) $S'(\mathcal{T}'(S(\mathcal{T}(I))))$. (e) Lazy wavelet transformed HDR image $S(I)$; (f) Tone-mapped (b) $\mathcal{T}(S(I))$; (g) Inverse Lazy wavelet transformed of $S'(\mathcal{T}(S(I)))$; (h) Inverse tone-mapped of (g) $\mathcal{T}'(S'(\mathcal{T}(S(I))))$ | 73 |
| 3.5 | The corresponding false colour map of 3.4. | 74 |
| 3.6 | 2D lifting structure of the proposed method: (a) Update first Forward lifting; (b) Update Inverse lifting; (c) Predict first Forward lifting; (d) Predict Inverse lifting. | 77 |
| 3.7 | (a) HH subband; (b) tone-mapped HH subband. | 88 |
| 3.8 | Results of Increasing Q_f of 5/3-p of EXR 007 (a) HDR-VDP, (b) MSE, (c) HDR-VDP of increased Q_f , (d) MSE of increased Q_f | 95 |
| 3.9 | (a) Original image of RGBE002, (b) original image of RGBE005. | 97 |
| 3.10 | HDR-VDP maps of RGBE002 ($Q_f = 256$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t | 97 |
| 3.11 | HDR-VDP maps of RGBE005 ($Q_f = 256$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t | 98 |
| 3.12 | (a) original image of RGBE006, (b) original image of RGBE017. | 98 |
| 3.13 | HDR-VDP maps of RGBE017 ($Q_f = 150$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t | 99 |
| 3.14 | HDR-VDP maps of RGBE009 ($Q_f = 150$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t | 99 |
| 3.15 | Results of HDR-VDP versus bit-rate (a) RGBE002, (b) RGBE005, (c) RGBE006, (d) RGBE017. | 100 |
| 3.16 | (a) Tone mapped No. 105 frame of Desert sequence; (b)Tone mapped No. 106 frame of Desert sequence. | 102 |

| | | |
|------|--|-----|
| 3.17 | (a) Temporal filtered Tone mapped No. 105 frame of Desert sequence; (b) Temporal filtered Tone mapped No. 106 frame of Desert sequence. | 102 |
| 4.1 | The usual HDR acquisition block diagram. | 106 |
| 4.2 | The existing HDR encoding block diagram. | 106 |
| 4.3 | The binary representation of RGBE format. | 106 |
| 4.4 | The binary representation of EXR format. | 107 |
| 4.5 | The proposed mapping-free HDR content compression block diagram. | 107 |
| 4.6 | Errors between single-precision floating-point and half-precision floating-point formats conversion. | 108 |
| 4.7 | (a)IEEE floating to integer conversion, (b)TMO function | 109 |
| 4.8 | (a)Histogram of HDR, (b)Histogram of Tone mapped HDR, (c)Histogram of IEEE half-floating integer. | 109 |
| 4.9 | (a) 10, 12, 16-bit PQ transform and IEEE floating to integer conversion | 111 |
| 4.10 | Result of HEVC codings. | 112 |
| 4.11 | Open-loop and closed-loop performance of JPEG2000-XT compression (a) The MSE results of RGBE 014 image; (b) The HDR-VDP results of RGBE 014 image | 113 |
| 4.12 | Rate-distortion curve of RGBE 001/ 008/ 014 images. (a) HDR-VDP result; (b) MSE resut; (c)HDR-VDP result; (d)MSE resut; (e)HDR-VDP result; (f)MSE resut. | 117 |
| 4.13 | Rate-distortion curve of EXR 002/ 006/ 008 images. (a) HDR-VDP result; (b) MSE resut; (c)HDR-VDP result; (d)MSE resut; (e)HDR-VDP result; (f)MSE resut. | 118 |
| 4.14 | Rate-distortion curve of PFM 008/ 013/ 018 images. (a) HDR-VDP result; (b) MSE resut; (c)HDR-VDP result; (d)MSE resut; (e)HDR-VDP result; (f)MSE resut. | 119 |

| | | |
|------|---|-----|
| 4.15 | Rate-distortion curve of test video sequence 01. (a) HDR-VDP P75 of each frame, (b) MSE of each frame, (c) mean of HDR-VDP P75, (d) mean of MSE. – HDR-VDP P75 represents the overall number of pixels which are beyond 75% of detection probability of HDR-VDP, and the lower of HDR-VDP P75 means better compression quality. | 122 |
| 4.16 | Rate-distortion curve of test video sequence 02. (a) HDR-VDP P75 of each frame, (b) MSE of each frame, (c) mean of HDR-VDP P75, (d) mean of MSE.. . . . | 123 |
| 4.17 | Mapping-free HEVC encoding Rate-distortion curve of test video sequence 01-09, from left to right and top to bottom. | 124 |
| 5.1 | The Architecture of the proposed model. | 128 |
| 5.2 | (a,b) Tone mapped under exposure image and corresponding saliency map, (c,d) Tone mapped proper exposure image and corresponding saliency map, (e,f) Tone mapped over exposure image and corresponding saliency maps. | 131 |
| 5.3 | Example of data set,(a) girl (b) penguin,(c) cheetah | 133 |
| 5.4 | Examples of a set of S_m , T_m , D_m | 137 |
| 5.5 | (a,b) The consecutive frames (1 and 2), (c,d) The ground truth, (e,f) The saliency map of the proposed spatial model. | 138 |
| 5.6 | (a,b) The temporal saliency map of IttiM model and the unpredicted information by the proposed spatial model, (c,d) The temporal saliency map of IttiF model and the unpredicted information by the proposed spatial model, (e,f) The temporal saliency map of Oaks model and the unpredicted information by the proposed spatial model, (g,h) The temporal saliency map of Dong model and the unpredicted information by the proposed spatial model. | 139 |

| | | |
|------|---|-----|
| 5.7 | (a,b) The consecutive frames (30 and 31), (c,d) The ground truth, (e,f) The saliency map of the proposed spatial model. . . | 140 |
| 5.8 | (a,b) The temporal saliency map of IttiM model and the unpredicted information by the proposed spatial model, (c,d) The temporal saliency map of IttiF model and the unpredicted information by the proposed spatial model, (e,f) The temporal saliency map of Oaks model and the unpredicted information by the proposed spatial model, (g,h) The temporal saliency map of Dong model and the unpredicted information by the proposed spatial model. | 141 |
| 5.9 | The ROC under different thresholds of binarization of No.14 image and right for No.33 image. | 142 |
| 5.10 | (a) HDR image; (b) luminance; (c) ground truth, (d) Weber's contrast, (e) normalized contrast, (f) DWT model of HDR, (g) DWT Weber's contrast model of LDR, (h) DWT normalized contrast model of LDR, (i) DWT model of LDR, (j) Dong's model, (k) CF model, (l) Itti's model of HDR, (m) proposed model. | 145 |
| 5.11 | Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 14 image and right for No 15 image. | 146 |
| 5.12 | (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 14 image and right for No 15 image. | 147 |
| 5.13 | Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 17 image and right for No 20 image. | 148 |
| 5.14 | (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 17 image and right for No 20 image. | 149 |

| | | |
|------|--|-----|
| 5.15 | Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 22 image and right for No 24 image. | 150 |
| 5.16 | (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 22 image and right for No 24 image. | 151 |
| 5.17 | Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 33 image and right for No 39 image. | 152 |
| 5.18 | (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 33 image and right for No 39 image. | 153 |
| 5.19 | (a) The saliency mask for segmentation, (b) Salient regions after segmentation, (c) Non-Salient regions after segmentation. (d) The summation of decoded Salient regions and Non-Salient regions | 154 |
| 5.20 | The decoding block diagram of the proposed model. | 154 |
| 5.21 | The original image, fixation density map, fixation points, k-means clustered Gaussian weighted map. from left to right, and top to bottom. | 156 |
| 5.22 | RD curves saliency guided compression. | 157 |
| 5.23 | Results of proposed HDR compression application, (a, b) Segmentation by threshold, (c, d) Decoded image of JPEG2K-HDR – the compressed file size are 607 KB, when compression ratio is 20, (e, f) Decoded image of Visual Saliency guided JPEG2K-HDR – the compressed file size are 363 KB, when compression ratio is 20 for salient region, and 100 for non-salient region, (g, h) Decoded image of JPEG2K-HDR – the compressed file size are 121 KB, when compression ratio is 100 | 158 |

| | | |
|------|--|-----|
| 5.24 | Results of proposed HDR compression application, (a, b) HDR-VDP map of decoded image of JPEG2K-HDR, (c, d) HDR-VDP map of decoded image of Visual Saliency guided JPEG2K-HDR, (e, f) HDR-VDP map of decoded image of JPEG2K-HDR for compression ratio 20. | 159 |
| 5.25 | Results of proposed HDR compression application, (a, b) Segmentation by threshold, (c, d) Decoded image of JPEG2K-HDR – the compressed file size are 607 KB, when compression ratio is 20, (e, f) Decoded image of Visual Saliency guided JPEG2K-HDR – the compressed file size are 363 KB, when compression ratio is 20 for salient region, and 100 for non-salient region, (g, h) Decoded image of JPEG2K-HDR – the compressed file size are 121 KB, when compression ratio is 100 | 160 |
| 5.26 | Results of proposed HDR compression application, (a, b) HDR-VDP map of decoded image of JPEG2K-HDR, (c, d) HDR-VDP map of decoded image of Visual Saliency guided JPEG2K-HDR, (e, f) HDR-VDP map of decoded image of JPEG2K-HDR for compression ratio 20. | 161 |
| 5.27 | The colourmap used for visualizing HDR-VDP metric. | 161 |
| 5.28 | MSE and HDR-VDP of saliency guided JPEG2000-Motion compression. | 163 |
| 5.29 | HDR-VQM of saliency guided HEVC HDR Mapping-free compression. | 163 |
| 5.30 | HDR-VDP and MSE of saliency guided HEVC HDR Mapping-free compression. | 163 |
| 5.31 | HDR-VQM per frame results of the orange lanes in Figure 5.29 which are the salient layers of QP0 and QP10. | 164 |
| A1 | RGBE images in data-set. | 189 |
| A2 | RGBE images in data-set. | 190 |

| | | |
|-----|---------------------------------------|-----|
| A3 | EXR images in data-set. | 191 |
| A4 | PFM images in data-set. | 192 |
| A5 | Saliency estimation data-set. | 193 |
| A6 | EXR Sequences in data-set. | 193 |
| A7 | RGBE 001. | 194 |
| A8 | RGBE 002. | 194 |
| A9 | RGBE 003. | 195 |
| A10 | RGBE 004. | 195 |
| A11 | RGBE 005. | 195 |
| A12 | RGBE 006. | 196 |
| A13 | RGBE 007. | 196 |
| A14 | RGBE 008. | 196 |
| A15 | RGBE 009. | 197 |
| A16 | RGBE 010. | 197 |
| A17 | RGBE 011. | 197 |
| A18 | RGBE 012. | 198 |
| A19 | RGBE 013. | 198 |
| A20 | RGBE 014. | 198 |
| A21 | RGBE 015. | 199 |
| A22 | RGBE 016. | 199 |
| A23 | RGBE 017. | 199 |
| A24 | RGBE 018. | 200 |
| A25 | RGBE 019. | 200 |
| A26 | RGBE 020. | 200 |
| A27 | RGBE 021. | 201 |
| A28 | RGBE 022. | 201 |
| A29 | RGBE 023. | 201 |
| A30 | RGBE 024. | 202 |
| A31 | RGBE 025. | 202 |
| A32 | EXR 001. | 203 |

| | |
|-----------------------|-----|
| A33 EXR 002. | 203 |
| A34 EXR 003. | 203 |
| A35 EXR 004. | 204 |
| A36 EXR 005. | 204 |
| A37 EXR 006. | 204 |
| A38 EXR 007. | 205 |
| A39 EXR 008. | 205 |
| A40 EXR 009. | 205 |
| A41 RGBE 001. | 206 |
| A42 RGBE 002. | 206 |
| A43 RGBE 003. | 207 |
| A44 RGBE 004. | 207 |
| A45 RGBE 005. | 207 |
| A46 RGBE 006. | 208 |
| A47 RGBE 007. | 208 |
| A48 RGBE 008. | 208 |
| A49 RGBE 009. | 209 |
| A50 RGBE 010. | 209 |
| A51 RGBE 011. | 209 |
| A52 RGBE 012. | 210 |
| A53 RGBE 013. | 210 |
| A54 RGBE 014. | 210 |
| A55 RGBE 015. | 211 |
| A56 RGBE 016. | 212 |
| A57 RGBE 017. | 212 |
| A58 RGBE 018. | 212 |
| A59 RGBE 019. | 213 |
| A60 RGBE 020. | 213 |
| A61 RGBE 021. | 213 |
| A62 RGBE 022. | 214 |

| | |
|-----------------------|-----|
| A63 RGBE 023. | 214 |
| A64 RGBE 024. | 214 |
| A65 RGBE 025. | 215 |
| A66 EXR 001. | 216 |
| A67 EXR 002. | 216 |
| A68 EXR 003. | 216 |
| A69 EXR 004. | 217 |
| A70 EXR 005. | 217 |
| A71 EXR 006. | 217 |
| A72 EXR 007. | 218 |
| A73 EXR 008. | 218 |
| A74 EXR 009. | 218 |
| A75 PFM 001. | 219 |
| A76 PFM 002. | 219 |
| A77 PFM 003. | 219 |
| A78 PFM 004. | 220 |
| A79 PFM 005. | 220 |
| A80 PFM 006. | 220 |
| A81 PFM 007. | 221 |
| A82 PFM 008. | 221 |
| A83 PFM 009. | 221 |
| A84 PFM 010. | 222 |
| A85 PFM 011. | 222 |
| A86 PFM 012. | 222 |
| A87 PFM 013. | 223 |
| A88 PFM 014. | 223 |
| A89 PFM 015. | 223 |
| A90 PFM 016. | 224 |
| A91 PFM 017. | 224 |
| A92 PFM 018. | 224 |

| | |
|----------------------|-----|
| A93 PFM 019. | 225 |
| A94 PFM 020. | 225 |

List of Symbols

| Symbols in Chapter 3 | Description |
|----------------------|---|
| x | discrete signal |
| P | prediction lifting |
| U | update lifting |
| a | approximation sample |
| b | detail sample |
| h | low-pass filter |
| g | high-pass filter |
| a | left-top element of a 2 by 2 block |
| b | right-top element of a 2 by 2 block |
| c | left-bot element of a 2 by 2 block |
| d | right-bot element of a 2 by 2 block |
| a' | left-top element of a 2 by 2 block after lifting |
| b' | right-top element of a 2 by 2 block after lifting |
| c' | left-bot element of a 2 by 2 block after lifting |
| d' | right-bot element of a 2 by 2 block after lifting |
| LL | LL Sub-band of 2D DWT |
| HL | HL Sub-band of 2D DWT |
| LH | LH Sub-band of 2D DWT |
| HH | HH Sub-band of 2D DWT |
| \mathcal{T} | Tone mapping operation |
| \mathcal{T}' | Inverse Tone mapping operation |

| Symbols in Chapter 3 | Description |
|----------------------|-----------------------------------|
| S | 2D lazy wavelet transform |
| S' | inverse 2D lazy wavelet transform |
| Symbols in Chapter 4 | Description |
| T | ransform function of ITU-R BT2020 |
| s | sign bit of IEEE-754 |
| e | exponent bit of IEEE-754 |
| m | mantissa bit of IEEE-754 |
| R | red color channel |
| G | green color channel |
| B | blue color channel |
| Y | luminance channel |

| Symbols in Chapter 5 | Description |
|----------------------|---|
| p | brightness parameter of TMO |
| I_h | HDR image |
| I_l | tone mapped LDR image |
| I | Intensity channel of Itti's model |
| O | Orientation channel of Itti's model |
| c | higher level of Gaussian pyramids |
| s | lower level of Gaussian pyramids |
| S | Spatial saliency map |
| T | Temporal saliency map |
| D | Unpredicted and under-predicted temporal saliency |
| C | Contrast Feature model |
| $w(\Delta t)$ | weight function related to exposure time t |
| SM | HDR weighted saliency map |

List of Acronyms

| Acronym | Description |
|---------|--|
| ADC | Analog to-Digital Converter |
| ASCII | American Standard Code for Information Interchange |
| AVC | Advanced Video Coding |
| CAM | Color Appearance Model |
| CC | Linear Cross Correlation |
| CCD | Charge-Coupled Device |
| CF | Contrast-Feature |
| CIE | International Commission on Illumination |
| CIF | Common Intermediate Format |
| CRF | Camera Response Function |
| CRT | Cathode Ray Tube |
| CSF | Contrast Sensitivity Function |
| CPU | Central Processing Unit |
| DCT | Discrete Cosine Transform |
| DRIIQA | Dynamic Range Independent Image Quality Assessment |
| DWT | Discrete Wavelet Transform |
| EWPSNR | Eye-Weighted Peak Signal Noise Ratio |
| GPU | Graphics Processing Units |
| HD | High Definition |
| HDR | High Dynamic Range |
| HEVC | High Efficiency Video Coding |

| Acronym | Description |
|---------|--|
| HVS | Human Visual System |
| IEEE | nstitute of Electrical and Electronics Engineers |
| ITMO | Inverse Tone Mapping Operation |
| ITU | International Telecommunication Union |
| JCT-VC | Joint Collaborative Team on Video Coding |
| JND | Just Noticeable Difference |
| JPEG | Joint Photographic Experts Group |
| KLD | Kullback-Leibler Divergence |
| LCD | Liquid Crystal Displays |
| LDR | Low Dynamic Range |
| LED | Light-Emitting Diode |
| MEPG | Moving Picture Experts Group |
| MSE | Mean Square Error |
| PFM | Portable FloatMap |
| PR | Perfect Reconstruction |
| PSNR | Peak Signal Noise Ratio |
| RLE | Run-Length Encoding |
| ROC | Receiver and Operating Characteristic |
| sRGB | standard Red Green Blue |
| TMO | Tone Mapping Operation |
| VDP | Visual Difference Predictor |
| VQM | Video Quality Measurement |

Chapter 1

Introduction

1.1 Motivation

Conventional digital images are usually represented by three bytes (24-bit) per pixel such as standard Red Green Blue (sRGB), retain only 2 orders of magnitude of the dynamic range. The Low Dynamic Range (LDR) encodings are usually called output-referred formats because they are compromised with the display range of a common monitor. The typical Cathode-Ray Tube (CRT) displays' and the common Liquid Crystal Displays' (LCD) operation range is limited by hardware restriction [107, 34]. In the real world, the visible light covers a vast dynamic range of the intensity. For example, the sun at 12 p.m (midday). could be 100 million times shiner than starlight [39] night. The Human Visual System (HVS) is capable of adapting to 10 orders of magnitude luminance levels and can perceive of about 5 orders of magnitude simultaneously [107, 34]. The High Dynamic Range (HDR) encodings are called scene-referred standard which is intended to represent the actual physical value of the light rather than exposed value. In HDR imaging, light can now be captured and fully utilized for various applications without the need to linearise the signal and to deal with clamped values. The very dark and bright regions of a given real-world scene can be captured and recorded simultaneously onto a digital image or a video, without compromise to under-exposed

and over-exposed regions.

The following Figure 1.1 shows that when (b) HDR solution offers a balance overall scene brightness preconception when LDR single exposure (a) contains under-exposed and over-exposed regions. The light source from outside in this particular example. The conventional single exposure of digital camera fails to capture the whole range of the given scene while the HDR solution is capable.



Figure 1.1: (a) A LDR stacks of a given scene in different exposure times and (from shorter to longer exposure times – from top-left to right-bottom) (b) a tone mapped HDR version.

However, the storage and transmission cost of the HDR solution is greater than LDR 8-bit representation. The most commonly used floating-point format of HDR content is double floating which requires 64 bit per colour channel and it is eight times of conventional LDR imaging. Hence the compression of HDR is mandatory. The aim of this thesis is to address challenges in compressing high-definition HDR images so that they can be efficiently stored and transmitted.

In existing HDR compression methods, the Tone Mapping Operators (TMOs) are performed before the decorrelation transforms *i.e.* Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT). It is because of the TMOs reduce the HDR incoming dynamic range to LDR so that the legacy encoder can be applied and the TMOs are not supposed to process the negative values *i.e.* negative coefficients of transforms. Hence, the applications operate

on the transformed frequency domain are compromised, *i.e.* watermarking and saliency estimation [98, 99, 9]. This challenge is addressed in Chapter 3. Since the Joint Photographic Experts Group 2000 (JPEG2000) codec supports up to 16-bit depth which is two times of the JPEG codec supported, the HDR dynamic range is reduced by an approximate logarithmic function, followed by 16-bit per colour channel in JPEG2000 compression by JPEG2000-HDR. These methods require binary to floating-point conversion and a mapping step before feeding into JPEG and JPEG2000 codecs. Moreover, the TMOs change the dynamic range of the HDR contents, and the quantization errors are affected accordingly. The quantization errors are then further increased and decreased correspondingly when decoding the HDR content. This challenge is addressed in Chapter 4. These HDR image compression approaches fail to compress non-salient region differently than the salient region, when HVS perceive them differently [107] especially for HDR content. This challenge is addressed in Chapter 5.

1.2 Research Questions

The main research question of this thesis is to address the research questions of HDR visual contents compression. To achieve this goal, a set of sub-questions is to be studied and addressed, as followed:

- The gap of the state-of-the-art HDR compressions and formats. The role of TMO in HDR compression. The possibility to use visual saliency as a HDR compression quantization guidance (addressed in Chapter 2).
- To propose a HDR compression scheme with DWT and TMO for better compression performance and LDR backward compatibility (addressed in Chapter 3).
- To propose a mapping-free HDR compression scheme without TMO for less computational complexity (addressed in Chapter 4).

- To propose a spatial saliency model for HDR visual content, and use it as guidance to compress HDR visual content with different quantization parameters (in Chapter 5).

1.3 Contributions

The main contributions of this thesis are:

- A DWT-TMO HDR compression exploits the reversibility of tone mapping and scalability of DWT in Chapter 3;
- A mapping-free post HDR image compression based on JPEG and JPEG-2000 standard codecs for current HDR image formats. This approach exploits the structure of HDR formats. It has an equivalent compression performance and the lowest computational complexity compared to the existing HDR lossy compressions (**50% lower than the state of the art counterparts**) in Chapter 4;
- A spatial saliency model for HDR visual content outperforms others by **10% for spatial visual prediction** task with **70% lower computational complexity** in Chapter 5;
- A HDR video saliency evaluation suggests more than **90% temporal saliency** is predicted by the proposed spatial model in Chapter 5;
- A saliency-guided compression scheme for HDR visual content, by applying different quantization factors according to the predicted saliency map (**30% bit-rate saving**) in Chapter 5.

1.4 Thesis Structure

The remaining thesis is structured in five chapters as follows:

- Chapter 2 **Literature Review**

This chapter identifies the gaps of the state-of-the-art HDR compressions and formats, understands the role of TMO in HDR compression, assesses the possibility to use visual saliency as a HDR compression quantization guidance.

- **Chapter 3 DWT with TMO built-in HDR Visual Content Compression**

This chapter investigates the relationship of Discrete Wavelet Transform (DWT) lifting steps and TMO. It proposes a novel approach to compress HDR image by JPEG2000 codec while backward compatible to LDR. This approach exploits the reversibility of tone mapping and scalability of DWT.

- **Chapter 4 Mapping-free HDR Visual Content Compression**

This chapter evaluates the importance of the TMO in the HDR compression. It proposes a mapping-free post HDR image compression based on JPEG and JPEG2000 standard codecs for current HDR image formats. The proposed method exploits the structure of HDR formats for low complexity compression.

- **Chapter 5 Saliency-Guided HDR Visual Content Compression**

This chapter addresses the shortcomings of the current HDR visual saliency models, and HDR visual saliency-guided compression is explored. It proposes a spatial saliency model for HDR visual content by combining the contrast feature and multi-exposure approach. Furthermore, the experimental results of this chapter suggest more than 90% temporal saliency is predicted by the proposed spatial model. The proposed saliency model can be used to guide HDR compression by applying different quantization factors according to the predicted saliency map.

- **Chapter 6 Conclusions**

Chapter 2

Background and Literature Review

HDR imaging is a revolutionary technique that uses the real physical values of the light. This chapter introduces fundamental concept of light, and the working pipeline from how to acquire, format, display, and evaluate the HDR content. The chapter is organised into seven sections as follows.

- Section 2.1 **Light, Human Visual Adaptation, and Colour Spaces**
- Section 2.2 **Generating and Displaying HDR Content**
- Section 2.3 **Encoding HDR content**
- Section 2.4 **Tone-Mapping HDR content**
- Section 2.5 **Evaluation of HDR content**
- Section 2.6 **Visual content Saliency Models**
- Section 2.7 **Evaluation Tools and Datasets**

The research questions raised after investigation of the state-of-the-art methods.

- How to integrate the backward compatibility to JPEG2000-HDR? – Section 2.3 lead to Chapter 3.

- To investigate the importance of the TMO in the HDR compression – Sections 2.3 2.4 lead to Chapter 4.
- What are the shortcomings of the current HDR visual saliency models? How to compress HDR according to the visual saliency? – Sections 2.5 and 2.6 leads to Chapter 5.

2.1 Light, Human Visual Adaptation, and Colour Spaces

It is essential to identify the similarities and differences of HDR and LDR visual content before actual work. This section introduces fundamental concepts of light quantities and the definition of contrast, the human visual adaptation and colour spaces. The LDR image records the values which are mapped to the output devices. The HDR contents represent the actual physical value of the light.

2.1.1 Light Quantities and the Contrast

It is crucial to understand visible light and units in HDR imaging because HDR image records the real physical values of light, from very dark values to very bright one *i.e.* $(10^{-3} - 10^6 \text{ m/cd}^2)$.

Visible light has a very narrow spectrum between $400 - 700\text{nm}$. The basic element of the light is Photons, and it follows the wave-particle duality [39]. Light propagates through the space interacting with media, and the light can be absorbed, refracted, reflected, and transmitted. The human eyes can perceive the travelling of the light. The different wavelengths are stimulated as colour visual sensations, see Figure 2.1.

Definitions of the measurement of the light and its units over time, space or angle are Radiometry and Photometry [107]. They represent physical units and the units of perceived brightness of human eyes respectively. A normal vision spectral sensitivity of human observer can be modeled by a single curve

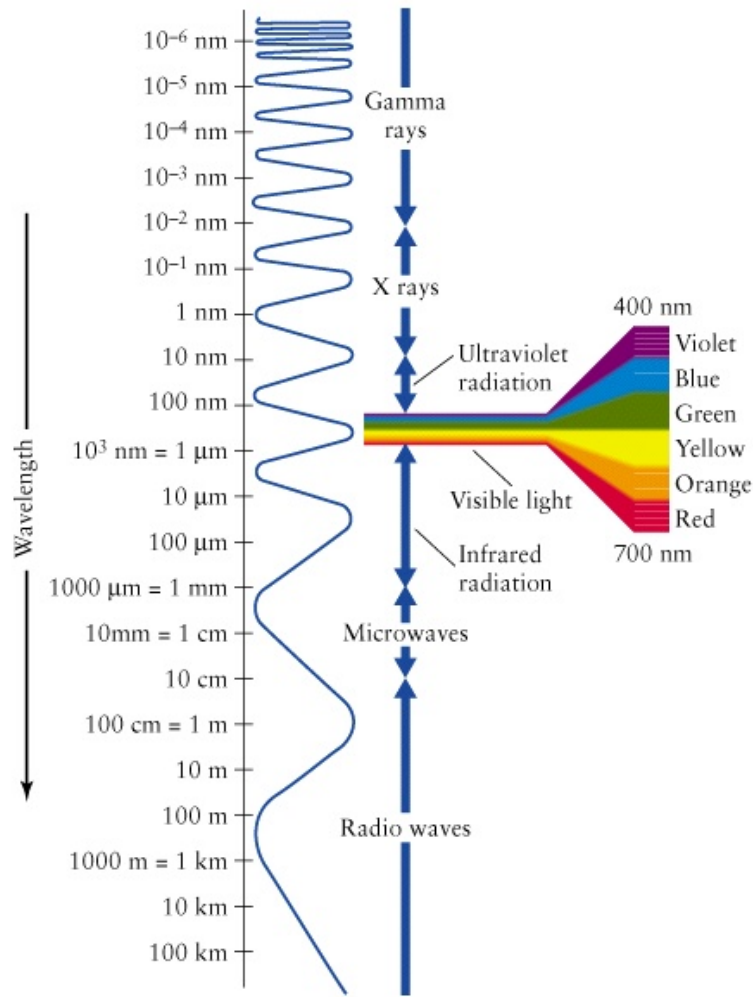


Figure 2.1: Wavelength image from [41]

see Figure 2.2. It is the Y component of the Commission on Illumination (CIE) XYZ space. These units and curve have followed the standard of the CIE.

The radiometric units:

1. Radiant Energy: the basic unit for light, joules (J). Q_e
2. Radiant Power: the amount of energy flows per unit of time ($J s^{-1}$).
3. Radiant Intensity: the amount of Radiant Power per unit of direction ($W sr^{-1}$).
4. Irradiance: the amount of Radiant Power per unit of area from all di-

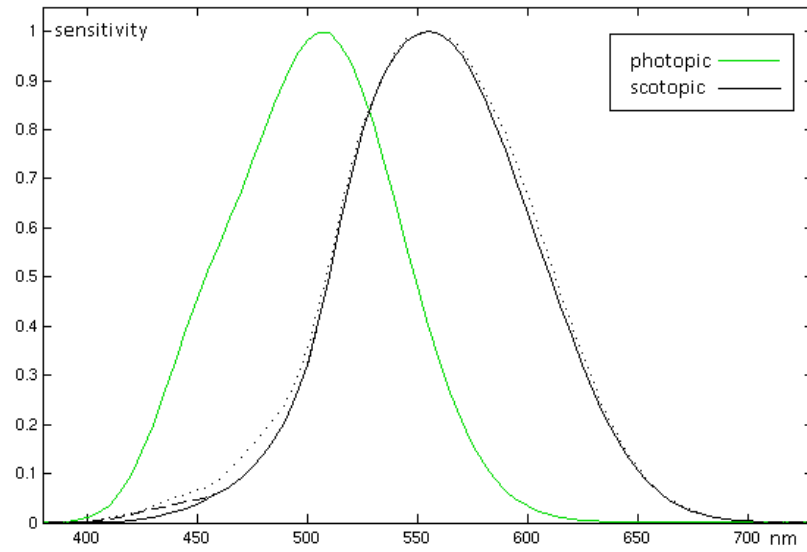


Figure 2.2: CIE standard observer photopic luminous efficiency curve from <http://www.cvrl.org>.

rections of the hemisphere at a point ($W m^{-2}$).

5. Radiance: the amount of Radiant Power trespassing at a point in a particular direction ($W m^{-2} sr^{-1}$).

The photometric units:

1. Luminous Power: the weighted Radiant Power, lumens is a derived unit from candela (P_v). lm .
2. Luminous Energy: the analogous of the Radiant Energy (Q_v). $lm s$.
3. Luminous Intensity: the Luminous Power per direction, candela (I_v). $lm sr^{-1} = cd$.
4. Illuminance: the analogous of the Irradiance (E_v). $cd m^{-2}$.
5. Luminance: the weighted Radiance (L_v). $cd m^{-2} = nit$

2.1.1.1 Contrast Ratio

Measurement of the relative luminance of the scene is very important which is called the contrast. It can be calculated by the minimum and maximum lumi-

nance values. There are 3 contrast measurement Weber contrast, Michelson contrast and Ratio contrast [102] are defined as:

$$C_W = \frac{L_{max} - L_{min}}{L_{min}}, \quad (2.1)$$

$$C_M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}, \quad (2.2)$$

$$C_R = \frac{L_{max}}{L_{min}}, \quad (2.3)$$

where L_{min} and L_{max} are the minimum and maximum luminance values of the scene and C_R is the ratio contrast definition which is the most adopted definition of the contrast.

2.1.2 Human Visual Adaptation

The HVS consisted of whole processes of visual perception. Real world's light is converted to electrical signals, then passed through the optical nerve to the visual cortex. The human eye could adapt huge dynamic range of light intensity. It is a fundamental component in the HDR imaging, the huge range of illumination in the single scene results in the different electrical signals by the pupil, the rod and cones, the photochemical reactions and photoreceptor mechanisms. In addition, the HVS perception of a scene depends on the lighting conditions.

Light, which passes through into the eye, firstly enters the Cornea. Then it comes into the Pupil, and the Iris responded in different sizes as apertures in a camera system. A refraction of light caused by the lens finally arrives the photoreceptors in the Retina. Then, the photoreceptors are categorized as cones and rods. The cones are sensitive to luminance levels from daylight to moonlight, and the perception of fast motion, high-frequency pattern. The perception of colour depends on the cones.

Three types of cones discriminate colour vision: cones for short wavelength are sensitive to wavelengths about $435nm$, cones for middle wavelength

are sensitive to wavelengths about $530nm$, and cones for long wavelength are sensitive to wavelengths about $580nm$.

The peak sensitivity of the rods is the luminance level of the night vision. The ability to distinguish colour is limited by only one type of rod. They are absent in the Fovea. Therefore, in the low light condition, the rods fail to discriminate the high-frequency patterns. The dynamic range of incoming light is reduced by the rods and cones. This sigmoid compression can be approximated by the following model:

$$\frac{P}{P_{Max}} = \frac{I^n}{I^n + \sigma^n}, \quad (2.4)$$

where P is the photoreceptor response, the light intensity is I . The n and σ are the the sensitivity control exponent and semi-saturation constant respectively [107, 92, 34].

2.1.3 Color Spaces

Color spaces are mathematical representations of colours. The conversion between colour vector and the standard CIE XYZ colour space, and the maximum and minimum primary values are determined by sets of equations. The colour vectors are usually as three components *e.g.* RGB, $Y_C B_C R_C$, CIE XYZ, LMS, $L\alpha\beta$, CIE Luv, CIE Lab and IPT [115] which are called primary colours. A colour space is usually modeling human perception and the capabilities to display devices which can be a camera, a LCD monitor, a CRT monitor or printed paper.

- **XYZ:** The CIE 1931 XYZ colour space is based on the response of short (S), middle (M) and long (L) wavelength rods' responses. The concept is that a colour sensation can be described as an additive model based on the amount of three primary colours ($S, M, and L$). XYZ is usually explained as a spectral power distribution I projection, into the

responses of rods or colour-matching functions, \bar{x} , \bar{y} , and \bar{z} :

$$\begin{aligned} X &= \int_{380}^{830} I(\lambda) \bar{x}(\lambda) d(\lambda), \\ Y &= \int_{380}^{830} I(\lambda) \bar{y}(\lambda) d(\lambda), \\ Z &= \int_{380}^{830} I(\lambda) \bar{z}(\lambda) d(\lambda), \end{aligned} \quad (2.5)$$

XYZ colour space was created for using the Y component to measure the luminance of the colour. The information relation for the hue and colourfulness of the colour or chromaticity is derived from XYZ values as:

$$\begin{aligned} x &= \frac{X}{X + Y + Z}, \\ y &= \frac{Y}{X + Y + Z}, \end{aligned} \quad (2.6)$$

- **RGB-L(luminance)** In HDR imaging, luminance channel usually provide the most useful information. Luminance is computed as:

$$L = Y = 0.2126R + 0.7152G + 0.0722B, \quad (2.7)$$

- **RGB-YC_BC_R** The YC_BC_R space is driven by the HVS is more sensitive luminance than chrominance. Hence the sub-sampling of chromatic channels in YC_BC_R space is useful in encoding applications. Conversion from RGB to YC_BC_R is given by:

$$\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.168 & -0.333 & 0.498 \\ 0.498 & -0.417 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (2.8)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.397 \\ 1 & -0.343 & -0.711 \\ 1 & 1.765 & 0 \end{bmatrix} \begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix}, \quad (2.9)$$

- **RGB-XYZ**

The white point and primaries determine the conversion matrix. Following equations are for conversion from RGB to XYZ of unknown white point cases:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (2.10)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2405 & -1.5371 & -0.4985 \\ -0.9693 & 1.8760 & 0.0416 \\ 0.0556 & -0.2040 & 1.0572 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (2.11)$$

- **XYZ-LMS**

The LMS colour space is derived from the responses of the three types of cones of the HVS, named by its sensitivity at long, medium and short wavelengths. Following equations are for conversion from XYZ to LMS of unknown white point cases:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.4002 & 0.7075 & -0.0807 \\ -0.2280 & 1.1500 & 0.0612 \\ 0 & 0 & 0.9184 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (2.12)$$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1.8502 & -1.1383 & 0.2384 \\ 0.3668 & 0.6439 & -0.0107 \\ 0 & 0 & 1.0889 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix}, \quad (2.13)$$

2.1.3.1 Gamma Correction

The relationship between the input voltage and output intensity of display device is non-linear. Hence, by applying the inverse power function value, the

values in RGB colour space are transformed to a non-linear input voltage:

$$\begin{bmatrix} R_v \\ G_v \\ B_v \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix}^{\frac{1}{\gamma}}, \quad (2.14)$$

where R_v, G_v, B_v are red, green, and blue channels' analog signal for display device, and γ is the value of power function.

The gamma correction is given by the above equations. It is very important to make sure samples were gamma corrected when doing a subjective comparison, and Figure 2.3 shows the comparison between the original and gamma 2.2 corrected images.

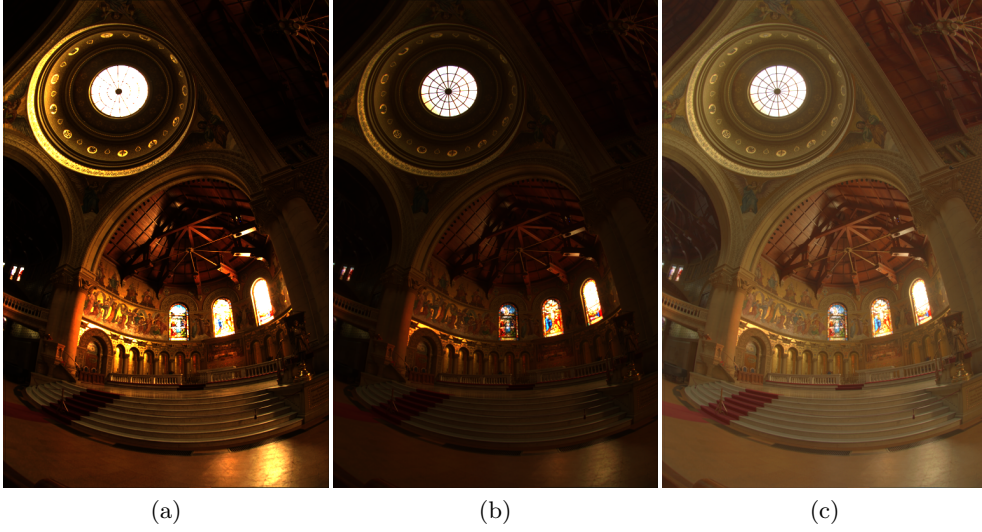


Figure 2.3: (a) HDR image; (b) TMO result; (c) Gamma corrected result.

2.2 Capturing and Displaying HDR Content

In this section, the methodologies and devices of generating and displaying HDR visual content are introduced.

2.2.1 Capturing of HDR Content

HDR content can be generated by computer graphics, photography, or both. The photography approach is time-consuming and equipment expensive. In the following sections, the intermediate HDR image and video capturing algorithms and the multi-exposure hardware solutions are introduced.

2.2.1.1 Recovering HDR Radiance maps from Photographs

The limitations of the image sensors restrict the dynamic range capturing ability of commercial cameras. The researchers proposed multiple exposures technique can overcome this problem by taking a stack of images and each pixel stores in the differently exposed values, see Figure 2.4. If the number of the stack and the exposure step size are enough, each pixel will be at least once properly exposed in the stack. Therefore, it is possible to recover the original light value of the scene and the full dynamic range [83].

However, most digital camera response curves (CRF) are not linear. Debevec *et al.* [24] proposed a method for recovering a CRF. They assumed that inverse CRF f is a monotonically increasing function f and irradiance E can be solved by linear optimization to get a CRF that minimizes the Mean Squared Error (MSE) over the derived CRF. The value of a pixel in the LDR image is given by a CRF and the exposure time:

$$I_k(x) = f(E_k(x)\Delta t_i), \quad (2.15)$$

where I is the intensity of image, k is the index of the colour channel, Δt_i is the exposure time. By applying logarithm to the equation:

$$\log(f^{-1}(I_k(x))) = \log E_{i,k}(x) + \log \Delta t_i \quad (2.16)$$

Then, f and E can be obtained by minimization of the least square error from:

$$O = \sum_{i=1}^N \sum_{j=1}^P \{w(Z_{ij})[g(Z_{ij}) - \ln E_i - \ln \Delta t_j]\}^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} [w(z)g''(z)]^2, \quad (2.17)$$

where i and j are the indices of pixel and exposure times, N is the number of pixel position and P is number of images of the stack. Z is the integer values of the pixel and λ is smoothing factor. The weighting function is given by:

$$w(z) = \begin{cases} z - Z_{min} & \text{for } z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z & \text{for } z > \frac{1}{2}(Z_{min} + Z_{max}) \end{cases}, \quad (2.18)$$

When the response curve g is recovered, the radiance values from pixel values are given by:

$$\ln E_i = g(Z_{ij}) - \ln \Delta t_j. \quad (2.19)$$

$$\ln E_i = \frac{\sum_{j=1}^P w(Z_{ij})(g(Z_{ij}) - \ln \Delta t_j)}{\sum_{j=1}^P w(Z_{ij})}, \quad (2.20)$$

An example is given in Figures 2.5 and 2.6.

2.2.1.2 Native Hardware Solutions

At the moment till the end of 2016 there is no currently available commercial 'HDR' camera. Since there is no single exposure solution that the sensor and Analog-to-Digital Converter (ADC) can losslessly capture and convert the whole dynamic range of the real world scene without noticeable noise, especially at the night. The most widely used solution is using at least a dual high-end camera simultaneously with a semitransparent mirror to shift the second camera exposure [42] by 8 to 16 times, in order to avoid ghosting artifact. The shortcoming is the base noise of low luminance scenario is obvious, the following examples in Figure 2.7 illustrate this mentioned effect. This is because to observe visible light in very low light conditions, the high light sensitive and it refers to the industry norm for sensitivity of emulsion based



Figure 2.4: (a) 1/250; (b) 1/60; (c) 1/15; (d) 1; (e) 4 ;(f) 15. secs of exposure times

film is used and the noise is increased in this case.

2.2.2 Display of HDR Content

At the moment, although so-called 'HDR' displays are advertised by the famous brands, their peak brightness could not compete with the HDR monitor for research purposes. The SIM2 HDR 47 display is the most widely used HDR monitor. It has HD 1080 resolution with a declared contrast ratio higher than

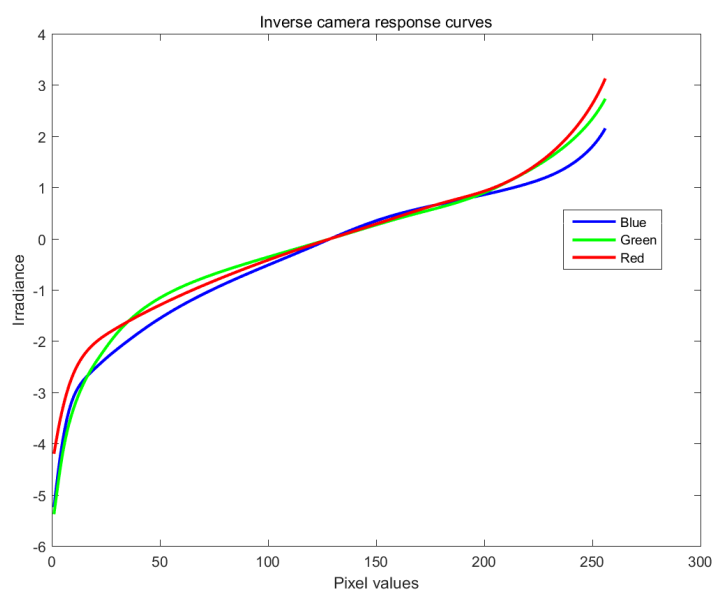


Figure 2.5: The inverse camera response curves.

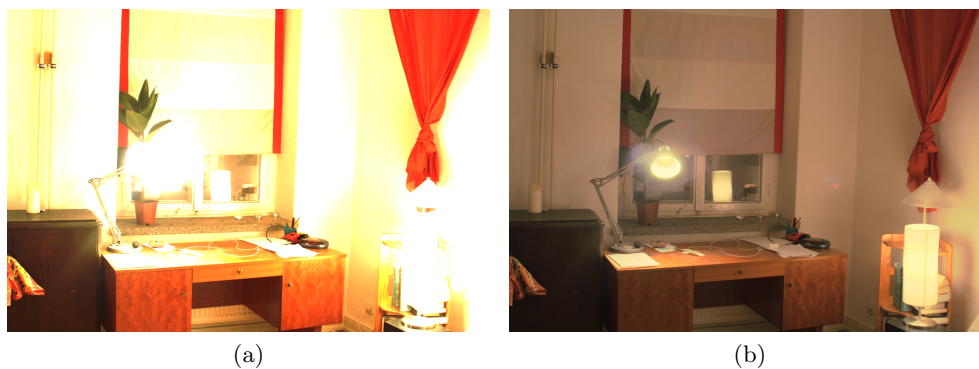


Figure 2.6: (a) recovered HDR; (b) tone-mapped.

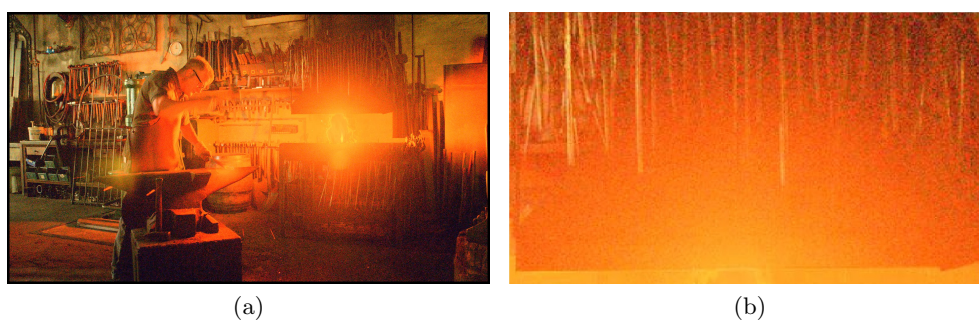


Figure 2.7: (a) A low light condition tone-mapped HDR frame; (b) zoomed-in version of (a).

$4 \cdot 10^6$. The peak luminance of SIM2 is around 4250cd/m^2 . The HDR display has an active matrix of Light-Emitting Diode (LED)s behind the LCD optical stack. This matrix of LED is equivalent to a low-resolution image that can be modulated in series with the LCD panel. The contrast of this LED image multiplies with the contrast of the LCD image which increases the dynamic range of the display. To minimize the light leaking from one LED to its adjacent areas and maximizing the contrast of the LED image, a reflector around the LED is introduced. The reflector spatially restricts the light from one LED to within a specified area surrounding the LED.

<http://hdr.sim2.it/>

2.3 Encoding HDR content

In the LDR imaging, the pixel values are usually 8-bit integer numbers. However, floating-point numbers are preferred to integers when encountering cross image addition and multiplication especially for values of the real world scene [1]. Therefore, the HDR values are generally recorded by single precision floating-point and it takes 96 bit per pixel. In contrast with the normal case of LDR imaging - 24 bit, which is 4 times of the amount of space required for an uncompressed LDR image.

In this section, the lossless and lossy HDR image encoding are introduced, then followed by HDR video compression.

2.3.1 HDR Image Lossless Formats

To address the amount of space requirement of HDR imaging and keep the fidelity of the real world scene, the HDR formats are proposed with lossless property with different error tolerances. The three most widely used lossless HDR image formats are: RGBE, EXR and PFM. The neighboured relative step size should be smaller than 1% to prevent visible luminance quantization effect according to [134]. Moreover, this is the reason why all relative step

Table 2.1: Established HDR image file formats.

| Format | Encodings | Compression |
|--------|--------------|--------------------------|
| HDR | RGBE XYZE | Run-length Run-length |
| EXR | Half RGB | Wavelet, ZIP |
| PFM | IEEE RGB | None Run-length |

sizes of HDR encodings satisfy such requirement for entropy encodings. The RGBE is the most commonly adopted HDR format of published papers and available online HDR sources, EXR follows next.

2.3.1.1 The Radiance RGBE/ XYZE

Ward [129] proposed RGBE (E represents the exponential component) file standard, and it contains a short American Standard Code for Information Interchange (ASCII) header. Which is a definition string of resolution of the image orientation and size. Then, the run-length encoded pixel data followed compose the file wrapper. The pixel data is either a 4-byte RGBE encoding or a CIE variant, XYZE, each channel stored in 8-bit (a byte). The dynamic range about HDR encodings is over 76 orders of magnitude, and the accuracy is beyond the Just Noticeable Difference (JND) of HVS. The Radiance files are encoded by run-length encoding with an approximately $\frac{1}{4}$ compression.

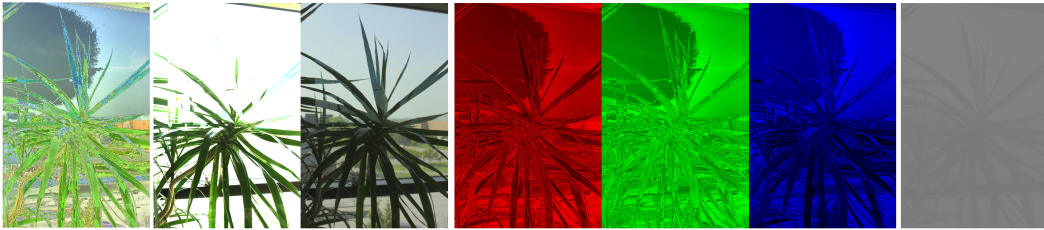


Figure 2.8: From left to right: HDR image reconstructed from only RGB channels of RGBE formats; HDR image reconstructed from RGBE channels of RGBE formats; Tone-mapped HDR image reconstructed from RGBE image; R; G; B; E channels of HDR RGBE image displayed separately

Figure 2.8 presents RGBE images displayed in R, G, B, E channels separately, because the exponent component is discontinuous heavily in R, G, B channels.

The following equations convert the scene-referred colour (R_w, G_w, B_w) to RGBE components (E, R_m, G_m, B_m) which is the procedure of HDR to RGBE encoding:

$$E = \log_2(\max(R_w, G_w, B_w)) + 128, \quad (2.21)$$

$$R_m = \frac{256R_w}{2^{E-128}}, \quad (2.22)$$

$$G_m = \frac{256G_w}{2^{E-128}}, \quad (2.23)$$

$$B_m = \frac{256B_w}{2^{E-128}}. \quad (2.24)$$

When an input in maximum (R, G, B) is less than 10^{-38} , the output is $(0, 0, 0, 0)$.

The inverse conversion from RGBE to HDR decoding is described as follows:

$$R_w = \frac{R_m + 0.5}{256} 2^{E-128}, \quad (2.25)$$

$$G_w = \frac{G_m + 0.5}{256} 2^{E-128}, \quad (2.26)$$

$$B_w = \frac{B_m + 0.5}{256} 2^{E-128}. \quad (2.27)$$

2.3.1.2 Logluv32

Ward proposed Logluv32 format[67], which is more perceptually motivated than RGBE by allocating more bit for luminance channel. The LogLuv32 HDR format is usually stored as a Tagged Image File Format (TIFF) file. Conversion between scene-referred pixel (r, g, b) and (X, Y, Z) is demonstrated by the following equations: Conversion of RGB to CIEXYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.212671 & 0.715160 & 0.072169 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (2.28)$$

Next is to define chrominance components.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \lfloor 410u' \rfloor \\ \lfloor 410v' \rfloor \end{bmatrix}, \quad (2.29)$$

the $\lfloor \cdot \rfloor$ is function of truncation,

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{1}{-2X + 12Y + 3} \begin{bmatrix} 4X \\ 9Y \end{bmatrix}. \quad (2.30)$$

Luminance Y is converted to logarithmic domain by:

$$L = \lfloor 256(\log_2 |Y| + 64) \rfloor, \quad (2.31)$$

the dynamic range of luminance is now recorded into 15-bit integer representation.

To recover rgb , XYZ is converted by:

$$Y = 2^{\frac{L+0.5}{256}-64}, \quad (2.32)$$

$$\begin{bmatrix} X \\ Z \end{bmatrix} = \frac{Y}{4v'} \begin{bmatrix} 9u' \\ -3u' - 20v' + 12 \end{bmatrix}, \quad (2.33)$$

where

$$\begin{bmatrix} u' \\ v' \end{bmatrix} = \frac{1}{410} \begin{bmatrix} u + 0.5 \\ v + 0.5 \end{bmatrix}, \quad (2.34)$$

L , u , v , and sign bit are encoded in 15, 8, 8 and 1 bit. Each channel is compressed by RLE independently.

2.3.1.3 EXR

EXR format[63] supports 16-bit half data type which is a logical contraction of the IEEE-754 [1] floating-point representation of 16-bit binary. Therefore,

EXR format requires 48-bit/pixel for the full colour format. EXR could record negative primary values. The dynamic range is about 10.7 orders of magnitude with a relative precision of 0.1% (10 times of RGBE) and the entire visible gamut [107].

Following formulas can convert 16-bit binary to EXR half precision presentation. When the exponent is in [1,30]:

$$I = (-1)^s * 2^{(e-15)} * 1.m, \quad (2.35)$$

when exponent - e is 0, and the mantissa - m is not 0:

$$I = (-1)^s * 2^{-14} * 0.m, \quad (2.36)$$

when exponent is 0, and the mantissa is 0:

$$I = 0, \quad (2.37)$$

when exponent is 31, and the mantissa is 0:

$$I = \infty, \quad (2.38)$$

when exponent is 31, and the mantissa is not 0:

$$I = NaN.-\text{not a number according to the IEEE-754 [1]} \quad (2.39)$$

2.3.1.4 PFM

PFM format records the HDR image by endianness order, and it is the default format of JPEG-XT standard [110, 111, 112]. The data type of PFM format is the IEEE 32 bit floating-point number similar to EXR format as described above.

2.3.2 Encoding HDR image Lossy HDR image Compression

Since lossless HDR image formats only applied entropy encoding, the redundancy of real world scene beyond the HVS could not be fully compressed. Therefore, lossy JPEG-HDR, JPEG2000-HDR and JPEG-Xt are proposed to incorporate high compression performance demands. The HDR lossy encoding scheme can be categorized into two groups: Single-Layer HDR Image Encoding, and Backward-Compatible HDR Image Encoding. The term backward-compatible means the encoding stream contains an LDR display version of HDR image which requires TMOs. The TMO will be introduced after HDR encodings section.

2.3.2.1 JPEG-HDR

Ward proposed JPEG-HDR [131] which is an HDR image compatible version of JPEG [127, 103] standard.

Figure 2.9 shows the block diagram of JPEG-HDR coding steps [131]: Firstly, an LDR version of input HDR image is created by TMO. Then, by using the HDR input to divide the LDR layer, the residual layer – ‘Ratio Map (RM)’ (the ratio of HDR and LDR which is the quotient) is generated. Next, the two layers pass through to the legacy JPEG encoder. The decoding of HDR image is multiplication of the residual layer and the LDR layer. Finally the LDR version of HDR image can be obtained by only decoding the JPEG compressed tone-mapped LDR layer. The actual HDR, LDR and a ratio map of them are displayed in Figure 2.10

Furthermore, because the limitation of HVS for detecting the high-frequency changes in luminance, the RM can be down-sampled to reduce the cost of recovering HDR image. However, down-sampling causes halos artifacts around high-frequency patterns, and a correction should be applied. This problem can be solved by pre-correction of differences between down-sampling and up-sampling of the RM image to obtain RM_d . Then, the original HDR image is divided by RI_d instead of RM , the result is a tone-mapped image with

corrections. Then, the RI_d is stored in 8-bit in the logarithmic domain. To further preserving colour information: a compression of the gamut and a new YC_bC_r encoding are employed. The gamut compression is done by a global desaturation by:

$$S = 1 - \frac{R, G, B}{Y}, \quad (2.40)$$

the desaturation of each colour channel is achieved by:

$$R_c = (1 - \alpha S^{\beta-1}) \cdot + \alpha S^{\beta-1} \cdot R, \quad (2.41)$$

where α is saturation factor, β determines the contrast of colour. Then the image is encoded in the modified YC_bC_r space. The original gamut of an HDR image can be recovered by:

$$R_o = \begin{cases} 1.055R^{0.42} - 0.055, & \text{if } R > 0.0031308 \\ 12.92R, & \text{if } |R| \leq 0.0031308, \\ -1.055(-R)^{0.42} + 0.055, & \text{if } R < 0.0031308 \end{cases} \quad (2.42)$$

the above equation is repeated for the all colour channels. The advantage of this method is LDR backward-compatible, the tone-mapped JPEG file can be displayed for normal LDR devices.

2.3.2.2 JPEG2000-HDR

Xu *et al.* proposed JPEG2000-HDR [135] which is a single layer encoding method. Figure 2.11 shows the block diagram of JPEG2000-HDR coding steps: Before JPEG2000 [122, 20] encoder applied, the HDR image is mapped by logarithmic function to reduce the HDR image dynamic range into 16-bit per channel depth. Then the compressed HDR image is passed to JPEG2000 encoder 16-bit depth mode.

The actual encoding process begins with the compression of the dynamic

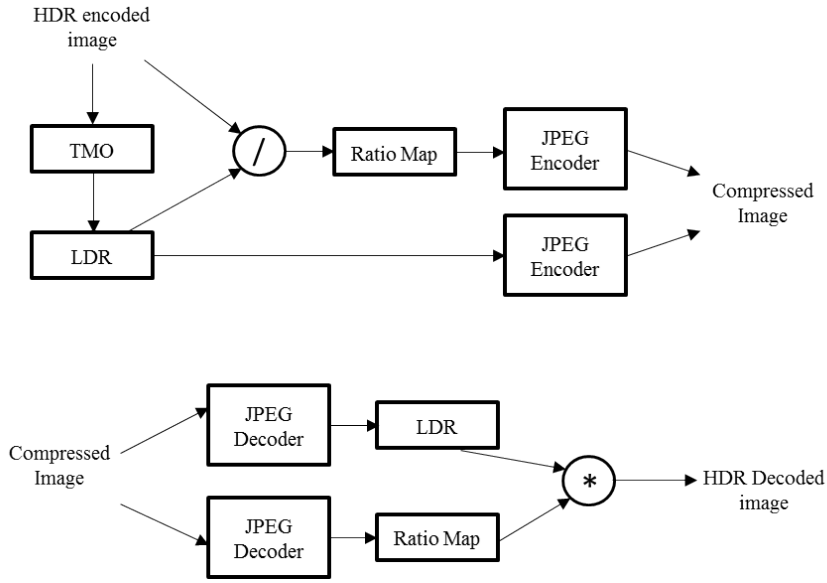


Figure 2.9: JPEG-HDR coding scheme

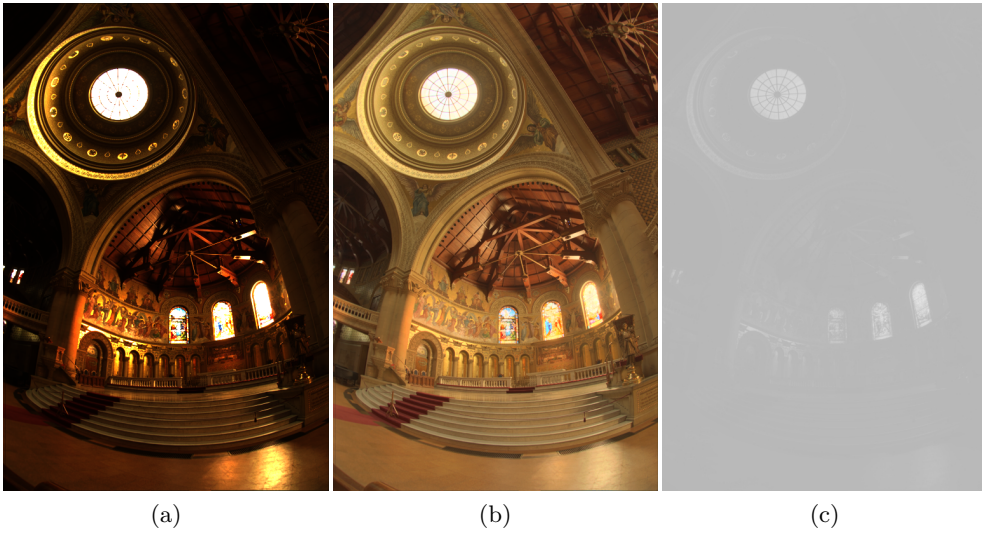


Figure 2.10: (a) HDR image; (b) LDR tone-mapped image; (c) Ratio map which is (a)/(b).

range applying the natural logarithm to the RGB values:

$$R_l = \log(R). \quad (2.43)$$

After that, the values of floating-point in the logarithmic domain are converted to 16-bit integers:

$$R_i = (2^{16-1}) \frac{R - R_{min}}{R_{max} - R_{min}}, \quad (2.44)$$

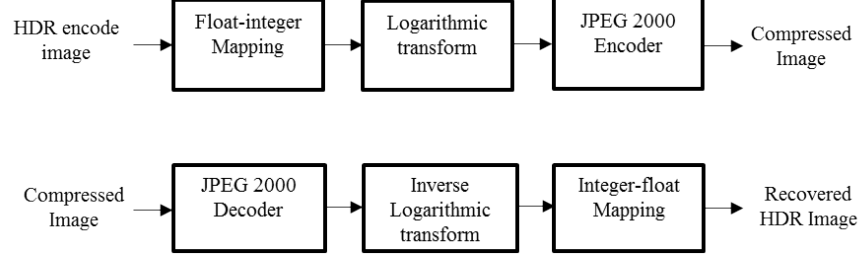


Figure 2.11: JPEG2000-HDR coding scheme

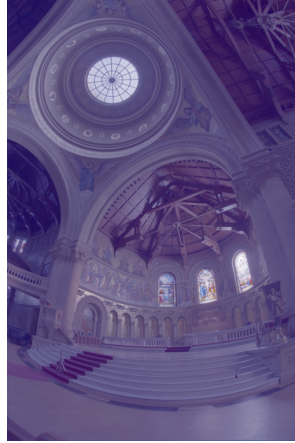


Figure 2.12: JPEG2000-HDR transformed integer image

The JPEG2000-HDR algorithm offers a straightforward approach of lossy compression for HDR images. Since the JPEG2000 codec is outperformed JPEG codecs, a better quality with less artifacts at low bit-rate can be realized in this method compared to JPEG-HDR method. The JPEG2000-HDR method dose not contains an individual LDR version of HDR image, the integer map of HDR image is displayed in Figure 2.12.

2.3.2.3 JPEG-XT

Richter [110, 111, 112] proposed JPEG-XT standard is an official JPEG extension, and it is backward-compatible.

Figure 2.13 shows the block diagram of JPEG-HDR profile-C coding steps: It is a closed-loop of two-layers encoding scheme, since JPEG supports only 8-bit per channel. The input image is tone-mapped and compressed first, then the compressed bitstream is decoded and inverse tone-mapped. The difference between the input image and the decoded and inverse tone-mapped bitstream is recoded as residual (refine) layer. Instead of division in the JPEG-HDR, the subtraction is used to record the difference between the original HDR image and the open-loop decoded HDR image.

The distortions of lossless mode and the highest setting of lossy mode of JPEG-XT increases dramatically which is worse than the JPEG2000 based methods.

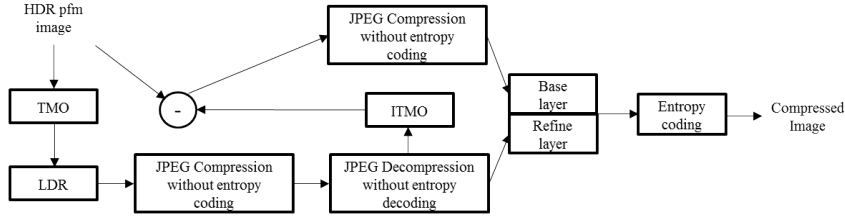


Figure 2.13: JPEG-XT coding scheme

2.3.2.4 Other HDR Compressions

There are some other compression schemes which support HDR encoding *i.e.* JPEG-XR; however, they are rarely used. Except these three most widely accepted approaches, there are other similar approaches:

JPEG-HDR type

1. Chen *et al.* [19] proposed JPEG compatible coding of HDR images. The bilateral filter and a detail layer are adopted instead of the remainder in the JPEG-HDR.

2. Kabaja [62] proposed the similar scheme with JPEG-HDR, the remainder image is encoded by wavelet encoder instead of JPEG encoder.

JPEG2000-HDR type

1. Shih *et al.* [114] proposed a similar method to JPEG2000-HDR, which has different colour transformation and floating-points to integer steps and it introduced JPEG2000 coding procedures into details.
2. Okuda *et al.* [100] proposed a method which replaced JPEG2000-HDR logarithmic transform and quantization steps with Logluv transform.
3. Sugiyama *et al.* [117] proposed a similar coding method to the JPEG2000-HDR, which replaced logarithmic transform and integer mapping by a TMO. The TMO can be adjusted by parameters, the parameters are optimized by minimizing the error between the HDR image and the LDR image.

Logluv type

1. Kikuchi *et al.* [64] proposed a lossless coding method which further compress Logluv format by a bit-plane coding.

2.3.3 Encoding HDR video

Existing HDR video coding schemes are based on well established video coding standards, *i.e.* H.264 [133, 91, 90, 113] and H.265 [118] for two main reasons:

1. Due to lacks of the development of HDR video hardware, standard and capture methodology, HDR video is rarely captured and shared.
2. Most commercial and civilian display devices are still LDR displays, there is no such strong need in developing a whole new video coding scheme especially for HDR videos.

2.3.3.1 HDR Video Compression based on AVC (MEPG4/ H.264)

There are three papers proposed by Reinhard and Yang based on H.264 codec. The bit-depth has been developed from 8-bit to 12-bit, and to 14-bit eventually.

Mantiuk *et al.* proposed a framework [87], identified and solved problem of losing high luminance sharp edge.

Then it is completed by a two-layers scheme [86] similar with JPEG-HDR with improvement by firstly passing through the CSF. Then, the invisible high-frequency redundancy was filtered by the wavelet transform, Contrast Sensitivity Function (CSF) and contrast threshold prediction, weighting function and inverse wavelet transform. When this method was proposed, the H.264 extension supports up to 12-bit luma encoding, however, it is rarely implemented in hardware, and Mantiuk *et al.* verified that 8-bit is enough for residual luminance frame with acceptable loss.

Zhang [142] proposed a new solution which has better results with additional 6 bit and reduce the redundancy by removing the invisible information of HVS [141, 140, 139]. Since Logluv only requires 16-bit for luminance, a conversion from 16-bit to 14-bit can achieve more accuracy compared to 16-bit to 8-bit. Therefore, the second layer (residual) for high-frequency detail can be avoided. At same bit-rate, it obtained 4/5 less of perceptual distortions than Logluv HDR video coding method [94].

2.3.3.2 HDR Video Compression based on HEVC (H.265)

Bits Reduction Transforms

At Joint Collaborative Team on Video Coding (JCT-VC) 16th meeting in Jan 2014, S.Lasserre *et al.* proposed JCTVC-P0159 HDR video coding based on H.265 (HEVC) [123]. They still adopted two layers coding scheme with 8-bit to 10-bit bit-depth. The novelty is it exploited the design of existing HDR displays, based on the modulation of an LDR picture by a LED backlighting. The Residual signal is considered as the backlighting signal for HDR displays. In their experiment, the extreme scenario occurred in the Seine 1080p 25fps

Table 2.2: HDR compressions

| | Core encoder | Single layer methods | Twolayers methods |
|-------|--------------|---|---|
| Image | JPEG | JPEG-XR [33] Li[73] | JPEG-HDR[131] JPEG-XT[111] Chen[19] Kabaja[62] |
| | JPEG2000 | JPEG2000-HDR[135] Shih[114] Okuda[100] sugiyama[117] | Masahiro[59] |
| | Indenpendant | Kikuchi [64] Zhang[142] Mai[81] | Jinno[60] |
| Video | MPEG | Mantiuk[86] | Mantiuk[87] |
| | HEVC | Dong[29, 7] Lasserre[123] | Lasserre[69] |
| | Indenpendant | Motra[94] Zhang[142] Mai[82] | |

sequence which recorded 21 f -stops, and the contrast ratio is about 2,000,000 :

1. Then, in the JCTVC-W15793, the 10 and 12-bit RGB to YCbCr transform of ITU-R BT 2020 standard was proposed.

HEVC-Range Extensions

The HEVC Range Extensions (RExt) [40] provide support for pixel bit-depths beyond 10-bit up to 16-bit [29, 7]. As a result, a half-precision 16-bit HDR compression without bit depth reduction is possible.

2.3.4 Discussion

HDR video and image compressions tend to utilize the existing coding standards (JPEG, JPEG2000, MPEG4, HEVC) with additional costs to record the HDR information (exponent, extra bit-depth), see Table 2.2. Most HDR encodings use TMO or logarithmic transform to reduce the dynamic range of HDR content, and the quantization error. Moreover, the reduced dynamic range content usually offers a LDR backward-compatibility. The shortcoming is the temporal artifacts will be introduced by applying TMO which affect the

visual perception. Furthermore, the existing methods are mostly high level compressions, which means the compression performances are determined by the core compression codes. *i.e.* The JPEG2000 based methods outperform JPEG based methods. The distortions of JPEG-Xt between quality setting 99 and 100 increase dramatically, *i.e.* from 1 to 0 HDR-VDP. The results of the Chapter4 suggested that the full bit-depth (*i.e.* 16) single-layer encodings outperform double-layer half bit-depth (*i.e.* 8) encodings. As a result, the support of higher bit-depth of the legacy codecs makes single layer encodings possible, and the TMO is not necessarily applied before compression.

More importantly, these HDR image compression approaches fail to compress non-salient region differently than the salient region, when HVS perceive them differently [107] especially for HDR content. The works of [101, 56, 49] suggested LDR content can be compressed by visual saliency. By extending this concept to HDR contents, the HDR compression schemes and visual quality evaluation can be more HVS driven. The visual saliency models are introduced in Section 2.6.

2.4 Tone-Mapping HDR Content

As mentioned in Section2.2.2, the HDR displays are rarely used and owned, and the LDR displays do not support the whole range of real world scene. Tone-mapping reduces the dynamic range of the input content to accommodate the dynamic range of the display devices. This operation has to keep some valuable properties of the original scene *i.e.* global and local contrast, details of edges. Moreover, the perception should be the same for both of the tone-mapped image and the real world scene. Usually, a TMO only maps luminance channel; the colours remain unchanged.

$$f(I) = L_d = f_L(L_w) \rightarrow [0 - 255], \quad (2.45)$$

$$\begin{bmatrix} R_d \\ G_d \\ B_d \end{bmatrix} = L_d \left(L_w^{-1} \begin{bmatrix} R_w \\ G_w \\ B_w \end{bmatrix} \right)^s, \quad (2.46)$$

where f is TMO, I is the HDR image, L_d is the LDR display luminance, L_w is the real world luminance, s is the saturation factor.

The log average luminance can be interpreted as the approximation of the scene, which indicates the overall brightness level. This L_w is computed by:

$$\bar{L}_{white} = \frac{1}{N} \exp \left(\sum_{x,y} (\log \delta + L_w(x, y)) \right), \quad (2.47)$$

where $L_w(x,y)$ is the real-world luminance for pixel (x,y), N is the total number of pixels in the image and δ is a small value to avoid the singularity that occurs if black pixels are present in the image.

The TMOs are categorized by their operation techniques T into different groups:

- Global Operators: the particular operation T is applied to every pixel.
- Local Operators: the mapping function T varies for each pixel, and it depends on their neighbouring pixels.
- Frequency/Gradient Operators: the mapping function T depends on different segmented regions of the image.
- Segmentation Operators: The operation separates the image into low and high frequencies layers, and the mapping T is applied to the low frequencies layer.
- Perceptual Operators: The mapping function T is corresponding to HVS characteristics
- Empirical Operators: The mapping function T is corresponding to photography theory

- Temporal Operators: The mapping function T is designed for tone-mapping HDR videos, T also minimises the artifacts occurred by spatial TMOs.

In this section, some of the state-of-the-art image and video TMOs with the artifacts are introduced.

2.4.1 Image TMO

Following subsections introduce classic HDR image TMOs with examples.¹ [24].

2.4.1.1 Brightness Reproduction

Tumblin and Rushmeier [125] proposed HVS driven brightness reproduction TMO. The TMO is defined as followed:

$$L_d = mL_{wa}L_{da}\left(\frac{L_w}{L_{wa}}\right)^{\gamma(L_{wa})/\gamma(L_{da})}, \quad (2.48)$$

where L_{da} and L_{wa} are display and image adaptation luminance respectively. γ is the human adapted to a luminance value function, which is defined as:

$$\gamma(L) = \begin{cases} 2.655, & \text{for } L > 100cd/m^2 \\ 1.855 + 0.4\lg_{10}(L + 2.3 \cdot 10^{-5}), & \text{otherwise} \end{cases} \quad (2.49)$$

m is the scaling of adaptation-dependent factor, the anomalous grey night images are avoided. It is defined as:

$$mL_{wa} = (\sqrt{C_{max}})^{\gamma_{wd}-1}, \quad (2.50)$$

The dynamic range of HDR image is compressed while the brightness is preserved. It generates plausible results when calibrated luminance values are

¹The example church HDR image for illustrating TMOs is courtesy of Paul Debevec

available.

The example shown in Figure 2.14 was generated by parameters: $L_{da} = 20$, $L_{dMax} = 100$, $C_{Max} = 100$, $L_{wa} = 0.0805$

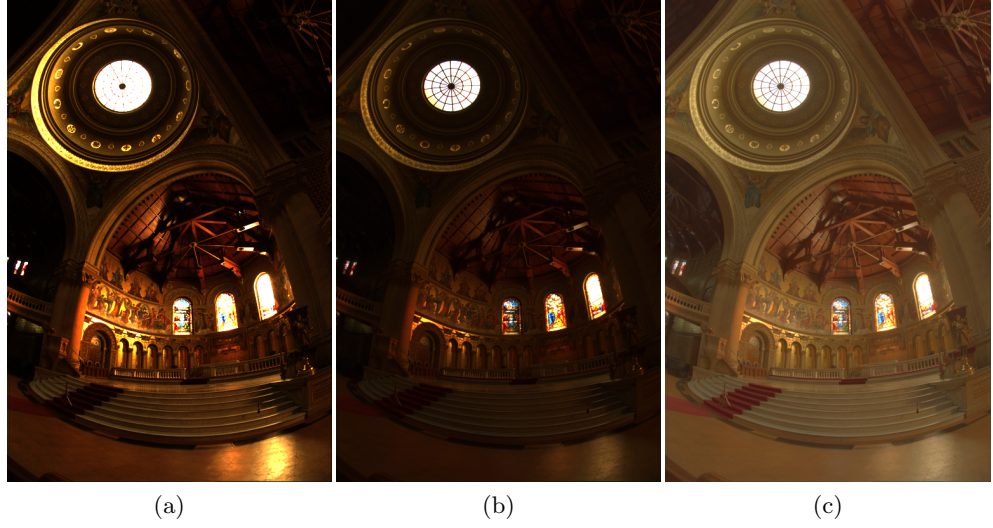


Figure 2.14: (a) HDR image; (b) TMO result; (c) Gamma corrected result.

2.4.1.2 Histogram Adjustment

Ward *et al.* [68] proposed TMO based on histogram equalisation method [47]. The glare, loss of acuity, and colour sensitivity are modeled by this method. The TMO is defined as followed:

The histogram h of the image in logarithmic domain \log is calculated first, using n numbers of bins. It is usually equal to 100. Then, the cumulative histogram H is computed as:

$$Hx = \sum_{i=1}^x h(b_i)/T \quad T = \sum_{i=1}^{n_{bin}} h(i), \quad (2.51)$$

where the histogram is its derivative with a proper scale:

$$\frac{dH(x)}{dx} = \frac{h(x)}{\sum_{i=1}^{n_{bin}} h(i)\Delta x} \quad \Delta x = \frac{1}{N} \log(L_{max}) - \log(L_{min}). \quad (2.52)$$

The histogram equalisation applied afterwards. *i.e.*

$$L_d \frac{f(\log(L_w))}{T \Delta x} \frac{\log(L_{d,max}) - \log(L_{d,min})}{L_w} = \frac{L_d}{L_w}. \quad (2.53)$$

The limitation of the contrast of HVS is modeled. The example shown in Figure 2.15.

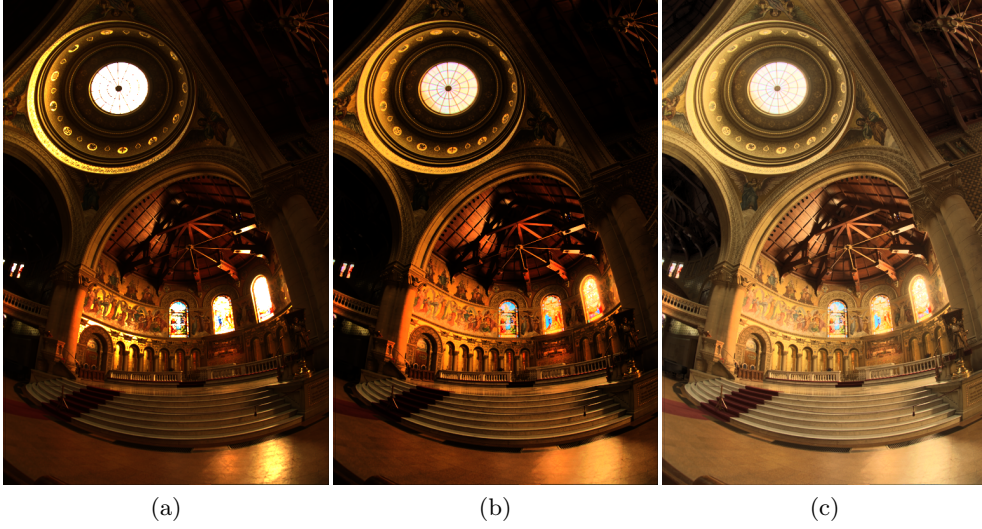


Figure 2.15: (a) HDR image; (b) TMO result; (c) Gamma corrected result.

2.4.1.3 Photographic Tone Reproduction

Reinhard *et al.* proposed photography inspired tone-mapping method [108].

It first approximates the exposure function in a real camera:

$$L(x, y) = \frac{\alpha}{L_w} L_w(x, y), \quad (2.54)$$

where $L(x, y)$ is luminance and α is brightness level of the scene.

Then, a S -shaped TMO is defined as:

$$L_d(x, y) = \frac{L(x, y)}{1 + L(x, y)}. \quad (2.55)$$

The photography TMO is defined by over exposure above equation:

$$L_d(x, y) = \frac{L(x, y)(1 + \frac{L(x, y)}{L_{white}^2})}{1 + L(x, y)}, \quad (2.56)$$

where L_{white} is the white point.

The local TMO can be realized by adopting different α based on the surroundings of each pixel. By applying centre-surround function based on the brightness perception model:

$$R_i(x, y, s) = \frac{1}{\pi(\alpha_i s)^2} \exp(-\frac{x^2 + y^2}{(\alpha_i s)^2}), \quad (2.57)$$

the response V_i is a convolution of:

$$V_i(x, y, s) = L(x, y) * R_i(x, y, s), \quad (2.58)$$

then by combining previous equations:

$$V(x, y, s) = \frac{V_1(x, y, s) - V_2(x, y, s)}{2^\phi a / s^2 + V_1(x, y, s)}, \quad (2.59)$$

Threshold V is introduced to minimize halo artifact with scale s_m :

$$|V(x, y, s_m)| < \epsilon. \quad (2.60)$$

where $\epsilon = 0.0001$ is threshold.

For a given pixel, $V_1(x, y, s_m)$ is approximation of the local average. The local TMO can be derived form the global:

$$L_d(x, y) = \frac{L(x, y)}{1 + V_1(x, y, s_m(x, y))}. \quad (2.61)$$

The examples of Reinhard global and local TMOs are shown in Figures 2.16 and 2.17 were generated by the parameters: $\alpha = 0.116$, $L_{White} = 306.89$, $\phi = 8$. The local contrast is well preserved in the local TMO in the windows and the shape of the dome especially.

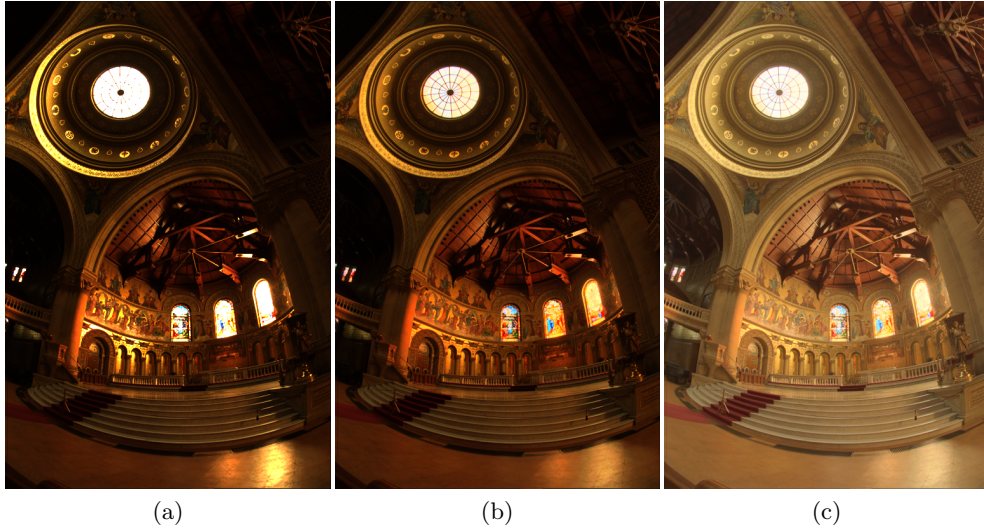


Figure 2.16: Reinhard global TMO (a) HDR image; (b) TMO result; (c) Gamma corrected result.

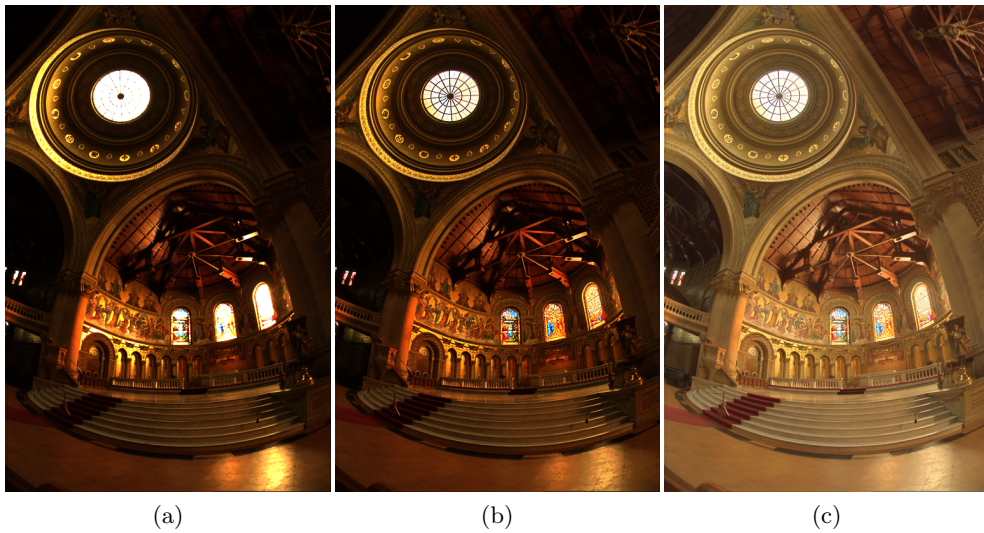


Figure 2.17: Reinhard local TMO(a) HDR image; (b) TMO result; (c) Gamma corrected result.

2.4.1.4 Adaptive Logarithmic Mapping

Drago *et al.* [32] proposed TMO based on logarithm. The TMO is defined as followed: The brightness, which is the response of the HVS to a luminance stimuli, presents a non-linear response. This can be approximated by using a logarithm, the Weber-Fechner Law, given by:

$$B = k_1 \log\left(\frac{L}{L_0}\right), \quad (2.62)$$

where L is the stimulus luminance, L_0 the backgrounds luminance, and k_1 is a constant. Following the Weber-Fechner Law, the mapping is defined as:

$$L_d = 10^{-2} \frac{L_{dmax}}{\log_{10}(L_{wmax} + 1)} \cdot \frac{\log(L_w + 1)}{\log(2 + 8(\frac{L_w}{L_{wmax}})^{\frac{\log(b)}{\log(0.5)}})}, \quad (2.63)$$

where b is a user parameter that controls contrast appearance L_{dmax} is used as a scale factor to adapt the output to its intended display.

The example shown in Figure 2.18 was generated by parameters: $L_{dMax} = 100$, $B = 0.85$.

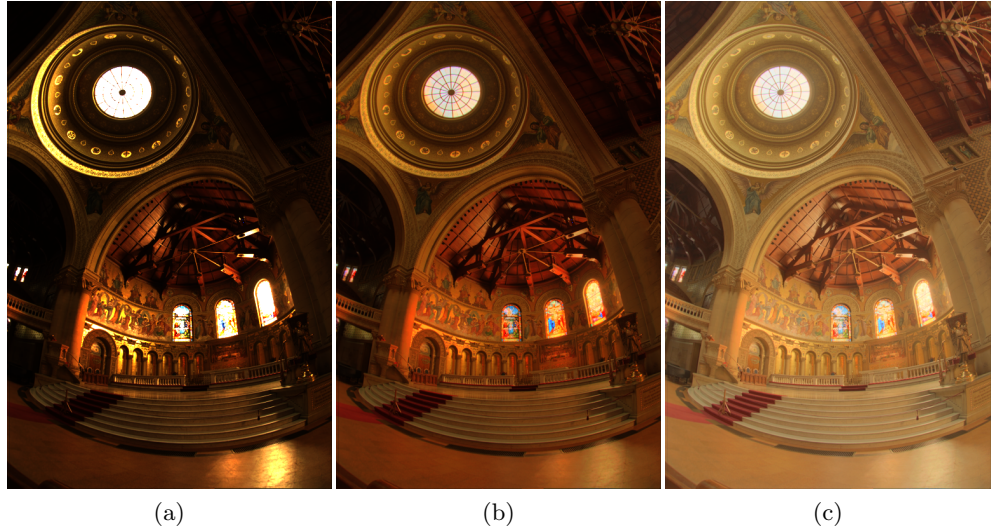


Figure 2.18: (a) HDR image; (b) TMO result; (c) Gamma corrected result.

2.4.1.5 Fast Bilateral Filtering

Durand and Dorsey [35] proposed TMO based on bilateral filtering. The image can be separated into a detail layer in high frequency and a base layer in low frequency by a bilateral filter.

The luminance of HDR image is calculated first. Then the bilateral filter is applied to the logarithmic domain of luminance. The detail layer is given by using the original HDR luminance to divide the filtered luminance. The TMO is applied to the filtered luminance. Finally, the tone-mapped base layer, the detail layer and the colour information are combined into a tone-mapped colour image.

The bilateral filtering is defined as:

$$D_j^s(x, y) = \frac{1}{k_j(x, y)} \sum_u \sum_v b_j(x, y, u, v) D(x - u, y - v), \quad (2.64)$$

$$k_j(x, y) = \sum_u \sum_v b_j(x, y, u, v), \quad (2.65)$$

$$b_j(x, y, u, v) = f(\sqrt{(x - u)^2 + (y - v)^2}) g(D(x - u, y - v) - D_j), \quad (2.66)$$

where D_j is a set of quantized possible values for given pixel (x, y) , g is Gaussian.

The example shown in Figure 2.19. It was generated by parameters: $\sigma_s = 0.5$, $\sigma_r = 4$ are the standard deviation of the space and range of g .

2.4.1.6 Computational Complexity Evaluation of TMO

In Table 2.3, the actual computational times of above TMOs are computed by a i5-4690K 3.5GHz central processing unit (CPU), 8GB random-access memory (RAM), MATLAB 2016b. The results are the average times per frame of the dataset used in this thesis.

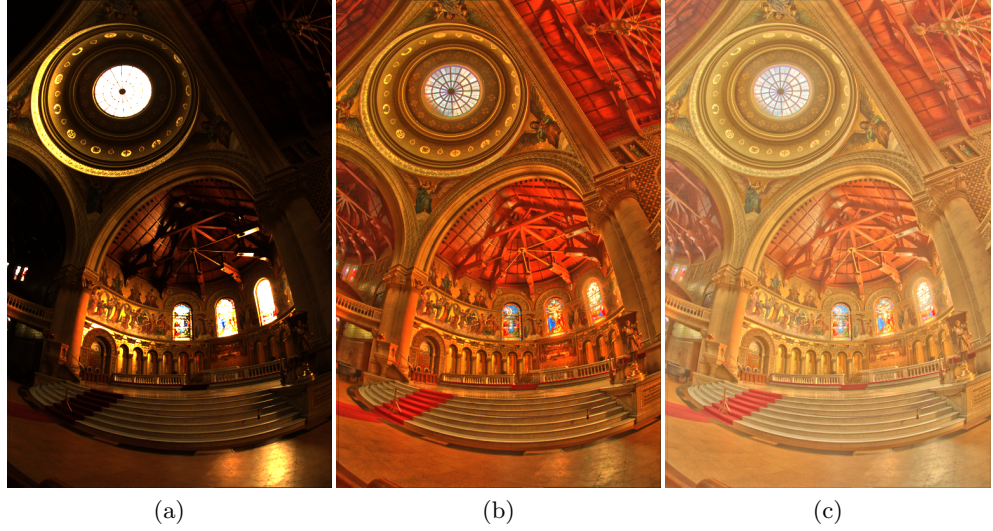


Figure 2.19: (a) HDR image; (b) TMO result; (c) Gamma corrected result.

Table 2.3: Computational time (secs per frame)

| TMOs | Computational Time |
|--------------------|--------------------|
| Tumblin | 0.17 |
| Ward | 0.16 |
| Reinhard Global | 0.32 |
| Reinhard Local | 2.07 |
| Drago | 0.22 |
| Durand | 0.43 |

2.4.2 Video TMO

Directly applying TMO to each frame for an HDR sequence generates temporal artifacts. Following subsections introduce the artifacts caused by TMO and recent HDR video TMOs to encompass those artifacts.

2.4.2.1 Video TMO artifacts

TMOs could cause three types of artifacts [14, 13]:

1. Flickering artifact(global and local)

The global flickering artifact is illustrated in Figure 2.20, the overall brightness suddenly changed due to TMO parameters.

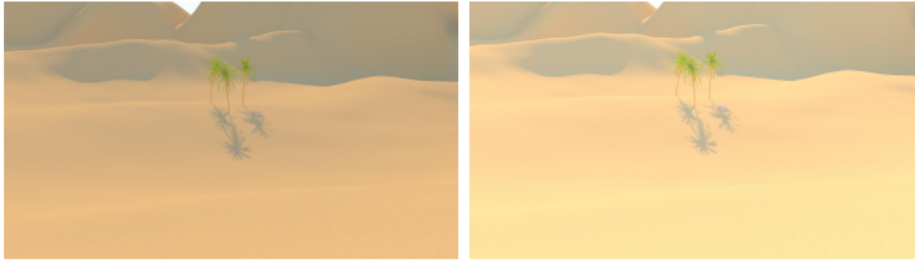


Figure 2.20: Example of global flickering artifacts



Figure 2.21: Example of local flickering artifacts

In Figure 2.21, a local flickering artifact is occurred due to bilateral filtering. A less filtered edges are in the middle frame compared to the neighbouring frames.

2. Temporal Contrast Incoherency

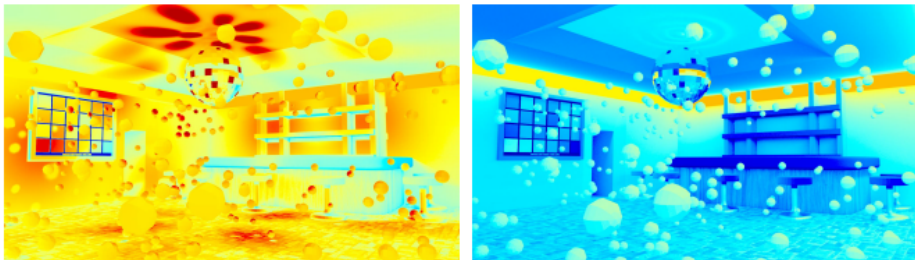


Figure 2.22: False colour luminance

Figures 2.22 and 2.23 give an example of temporal contrast incoherency in false colour due to lighting condition changing and TMO.

3. Temporal Object Incoherency

The temporal object incoherency is introduced in Figure 2.24, the levels of brightness of the bottom of the bridge are the same for HDR raw

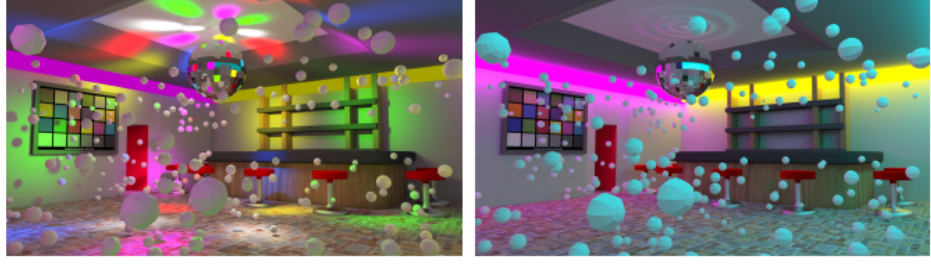


Figure 2.23: Temporal contrast incoherency occurs when illumination changed

sequence. However, after a TMO applied in Figure 2.25, the brightness

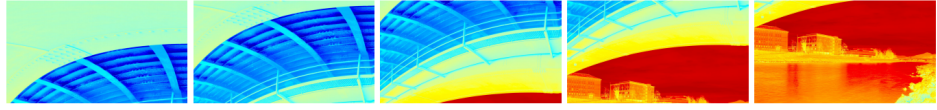


Figure 2.24: False colour luminance

of the bridge changed in tone-mapped sequence.



Figure 2.25: TMO applied

2.4.2.2 Temporally Coherent Local TMO

Aydin *et al.* [6] proposed a spatial filtering process to ensure high reproduction of contrast when ghost-free pixelwise temporal filtering is performed. It first decomposed every frame by base and detail layers with permeability map. Then, the motion flow is used for motion compensation. Next, the temporal filtering is performed for the both base and detail layers. A combination of tone-mapped base layer multiplied by detail layer to generate the Tone-mapped frame followed. The method relies on the interaction of the user to achieve the best trade-off between temporal and spatial contrast. Hence it is not suitable for automatic tasks.

2.4.2.3 Real-Time Noise-Aware TMO

Eilertsen *et al.* [36] proposed an improved bilateral filter model [35] for the separation of base and detail layers. The noise is suppressed by the de-noising filter. Also, the temporal artifacts are reduced by the low-pass filter, as suggested by Mantiuk *et al.* [84]

2.4.3 Discussion

Global TMOs are less computational complexity than local TMOs. However, global operations could not preserve the fine details. The visual perception of TMO depends on the parameters of the TMO; it is not suitable to use the same parameters for every HDR image. There are researches on TMO parameters estimation and toning [106, 46, 44]. The temporal artifacts of tone-mapped HDR videos are generated by the spatial TMOs. The temporal filtering of brightness of TMOs reduces these artifacts with the sacrifices of the overall brightness. In this thesis, the Reinhard global TMO is selected as example TMO because it has limited number of parameters which can be automatically estimated. function and obtained balance between computational complexity and pleasing perception. Moreover, the corresponding inverse function can be acquired analytically which is important for our compression application. The results of [31] shows that the Reinhard TMO mapping is the closest TMO to the ideal point extracted from preferences of participants, even though the sample size is low (11 participants and 4 image), the subjective ground truth results of [137] further validated the Reinhard TMO operates very similar to the HVS.

Moreover, the TMOs change the dynamic range of the HDR contents, and the quantization errors are affected accordingly. The quantization errors are then further increased and decreased correspondingly when decoding the HDR content.

2.5 Visual Quality Evaluation of HDR Content

Current image fidelity measurements mostly assume the same dynamic range of both reference and distortion images, and this range usually refers to the $0 - 255$. Although some MSE based existing quality assessments can be extended to the HDR, the main disadvantage is the lack of HVS model. It is crucial to utilize HVS for discriminating distortions. For example, certain distortion beyond human eyes observing capability can be neglected during the comparison. In this section, the quality metrics specialized in the HDR image are introduced.

2.5.1 Objective evaluation of HDR content

The HDR specialized metrics can deal with the HDR values [4]. These are:

- Root Mean Square Error (RMSE)
- multi-exposure Peak Signal Noise Ratio (mPSNR) [95]
- HDR Visual Difference Predictor (HDR-VDP) [88]
- Dynamic Range Independent Image Quality Assessment (DRIIQA) [5]
- High dynamic range Video Quality Measurement (HDR-VQM) [96]

The first two metrics are statistical metrics sensitive to relative small errors. For example, if an image and its shifted copy (few pixels) are compared, the error will be very high with these metrics. However, these images might be perceived the same. For Inverse Tone Mapping Operators (ITMO), where the reconstructed values are not the same as the original image, the pixel values are increased by a few intensity units. Therefore, the HDR perceptual metrics should be employed. In particular, the HDR-VDP is an extended version for the HDR images of the VDP. The VDP modeled limitations of the HVS instead of errors in the physical value domain for discrimination. The final results of HDR-VDP is a map of detection probability of perceptual difference between images.

2.5.1.1 NON-HVS Driven Metrics

Non-HVS driven metrics are usually computationally fast. However, the accuracy is not guaranteed. Hence, they are useful for the particular applications. The relatively small differences in very dark and very bright areas result differently, and the non-HVS driven metrics treat these differences the same. Furthermore, the peak of the signal in the HDR content is not pre-determined, and it is not fixed as certain values *i.e.* 255. The related non-HVS driven metrics are listed below.

- Mean squared error

The MSE is defined as:

$$MSE = \frac{1}{X * Y} \sum_{m=0}^{X-1} \sum_{n=0}^{Y-1} (I_1(m, n) - I_2(m, n))^2. \quad (2.67)$$

- Root mean squared error

The RMSE is defined as:

$$RMSE = \sqrt{MSE}. \quad (2.68)$$

- Peak signal-to-noise ratio

The PSNR is defined as:

$$PSNR = 10 \log_{10} \left(\frac{P^2}{MSE} \right) dB, \quad (2.69)$$

where P is the peak value of given image I .

- Multi exposure Peak signal-to-noise ratio

The mPSNR is defined as [95]:

$$mPSNR = 10 \log_{10} \left(\frac{3 * 255^2}{MSE'} \right), \quad (2.70)$$

where

$$MSE' = \frac{1}{n * w * h} \sum_c \sum_{x,y}^{w,h} (\Delta R_{x,y}^2 + \Delta G_{x,y}^2 + \Delta B_{x,y}^2), \quad (2.71)$$

where n is the number of exposures c , and w, h is the image resolution,

$$T(I) = \left[255(2^c I)^{1/\gamma} \right]_0^{255}, \quad (2.72)$$

where c is the exposure compensation in f-stops usually $(-8 \ 8)$, γ is the display gamma, and $[\]_0^{255}$ indicates clamping to the integer interval $[0, 255]$.

- Signal-to-noise ratio

The SNR is defined as:

$$SNR = 10 \log_{10} \left(\frac{\sum_{m=0}^{X-1} I_1(m, n)^2}{\sum_{m=0}^{X-1} \sum_{n=0}^{Y-1} (I_1(m, n) - I_2(m, n))^2} \right). \quad (2.73)$$

- Mean signal-to-noise ratio

The MRSE is defined as:

$$MRSE = -10 \log_{10} \left(\frac{\sum_{m=0}^{X-1} \sum_{n=0}^{Y-1} (I_1(m, n) - I_2(m, n))^2}{\sum_{m=0}^{X-1} \sum_{n=0}^{Y-1} (I_1(m, n)^2 + I_2(m, n)^2)} \right). \quad (2.74)$$

2.5.1.2 Visible Differences Predictor

Daly [21] proposed VDP which is an assessment of image fidelity. It uses CSF to describe the visual ability of seeing objects that may not be outlined clearly or that do not stand out from their background. In Figure 2.26 the frequency of visual pattern increases from left to right with the narrower pattern. From bottom to top, the pattern decreases in contrast with a lighter pattern. The boundary between visible and invisible patterns determines the CSF curve shape.



Figure 2.26: Sweep grating to measure CSF

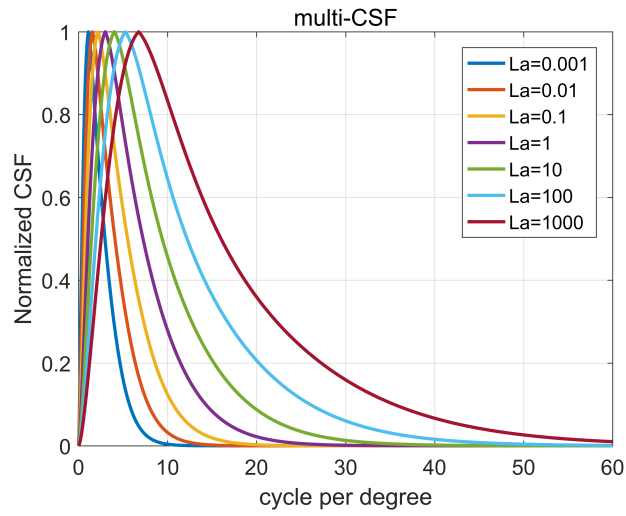


Figure 2.27: CSF under different adaptation levels

According to the Dely's experiment, the CSF is given by:

$$CSF(\rho, \theta, L_a, i^2, d, c) = P \times \min[S_1(\frac{\rho}{r_a \times r_c \times r_\theta}), S_1(\rho)], \quad (2.75)$$

where

$$\begin{aligned} r_a &= 0.856 \times d^{0.14}, \\ r_c &= \frac{1}{1 + 0.24c}, \\ r_\theta &= 0.11 \cos(4\theta) + 0.89. \end{aligned} \tag{2.76}$$

$$\begin{aligned} S_1(\rho) &= [(3.23(\rho^2 i^2)^{-0.3})^5 + 1]^{-0.2} \times A_{l\epsilon\rho}^{(-B_{l\epsilon\rho})} \sqrt{1 + 0.06e^{B_{l\epsilon\rho}}}, \\ A_l &= 0.801(1 + 0.7L_a^{-1})^{-0.2}, \\ B_l &= 0.3(1 + 100L_a^{-1})^{0.15}. \end{aligned} \tag{2.77}$$

where ρ is spatial frequency in cycles per visual degree, θ -orientation ([18] [132]), the θ should be stimuli's orientation, L_a - the light adaptation level in cd/m^2 , i -the stimulus size in deg^2 , d -distance in meters, c -eccentricity, ϵ -constant=0.9, and P -the absolute peak sensitivity = $250cd/m^2$ [5].

2.5.1.3 Dynamic Range Independent Image Quality Assessment (DRIIQA)

Aydin *et al.* [5] proposed a metric which combines structural similarity metric (SSIM) and an additional contrast detection part. This metric detects visible contrast changes and classifies them into three kinds of structural change: loss of visible contrast, amplification of invisible contrast, reversal of visible contrast. A unique feature is able to compare images in different dynamic ranges. Therefore, this metric can evaluate TMO and ITMO objectively rather than relying on the observer's subjective judgement.

2.5.1.4 HDR-VDP2

HDR-VDP [88] is a strict HDR visual quality assessment. The discriminate of HDR distortion ability is better than SSIM metric.

HDR-VDP metric is based on the modeling of HVS by including CSF; the intra-ocular light scatters; photoreceptor spectral sensitivities; separation of rod and cone pathways; contrast masking of intra and inter channels.

Discussion: It is worth mentioning that both HDR-VDP and DRIIQA only

compute differences in luminance channel but not in full colour, and most of the TMOs have the same intention. To evaluate the proposed HDR compression scheme, the HDR-VDP is more appropriate because the error due to the compression sometimes are not affecting the perception, and the dynamic range is unchanged.

2.5.1.5 HDR-VQM

Narwaria [96] proposed an HDR video quality measurement. It first converts the input luminance to displayed luminance. Next, the effect of the distortions is analysed by comparing of the different frequencies and orientation sub-bands. Finally, the error pooling is a combination of short-term temporal and spatial pooling for the final video quality score. However, the ability to discriminate the distortion of this metric is not as good as HDR-VDP metric by the experimental results in Chapter 5.

2.5.2 Subjective Evaluation of HDR Content

The objective visual metrics are designed to match the subjective perceptual experience. Therefore, the well-designed subjective visual evaluation can be seen as the ground truth for objective visual metrics models [109]. The reference stimulus usually are the original HDR visual contents, the comparison stimulus are usually compressed, distorted or tone-mapped HDR or LDR visual content [54, 53]. There are two common ways to subjectively evaluate HDR visual content

Rating: a participant has to rate the stimulus within a scale.

Pairwise Comparisons: a participant has to determine whether one out of a pair image is closer to a reference.

There is no subjective evaluation in this thesis, because the HDR displays are expensive.

2.5.3 Discussion

Due to the difficulties of HDR subjective quality evaluation, the objective metrics are used to evaluate the distortion of proposed compression scheme.

The non-HVS driven metrics provide statistical measurement of relative distortion between two images, the processes are simple and the results are robust for different viewing conditions, and less accurate for particular viewing condition. In contrast, the HVS driven quality metrics modeled by HVS aspect, the processes are complex and the results are estimated for particular viewing condition. Moreover, the HVS driven metrics are usually use pooling rules for pixel-wise, the large pixel-wise error value is amplified. The experimental results suggest HDR-VQM metric fails to differentiate small distortion, *i.e.* HEVC QP0 and QP10 encoded contents in Chapter 5. The HDR-VDP can detect this difference. The DRIIQA metric focuses on the HVS contrast perception of the scene instead of distortions, *i.e.* it is more sensitive to the brightness and contrast changes, and fails to detect the Gaussian noise. In this thesis, the HDR-VDP and MSE used as main quality evaluation metric to evaluate the compression distortions.

The above visual quality metrics fail to considering the distortions differently according to visual saliency. The HVS driven metric tend to weigh mid-frequency error differently, by not in the particular visual saliency approach. The eye-weighted PNSR can be extended to these objective metrics, however, it requires the actual eye-fixation points from eye-tracker.

The objective evaluation metrics provide an inexpensive way for analysing HDR visual content compared to subjective evaluation. However, the objective metrics are approximations. Hence, the accuracy is not guaranteed. The results in [51] illustrate that the HDR-VDP metric is the closest visual quality metric to the subjective ground truth HDR.

2.6 Visual Saliency Models

Visual saliency represents the unique properties of a particular scene. It is usually referred to objects or areas that stimulate an observer more than surrounding neighbouring parts. The term 'salient' is usually interpreted as in the context of bottom-up models. Many LDR approaches with HVS based assumptions for visual attention algorithms have been proposed and evaluated in the related researches.

In the HDR imaging, the saliency detection models are useful for many applications, *i.e.* TMO: local TMO for the salient regions, global TMO for the non-salient regions; Compression: high quality for the salient regions, low quality for non-salient regions.

2.6.1 Spatial Saliency Models

In the real world, visual attention usually constantly changes due to egocentric movements in the scene. Then visual selection depends on the both current scene saliency and the prior knowledge of past scene. Hence, the visual attention model is designed for capturing regions of the scene that are more relatively important in a spatio-temporal search. In the next subsections, the classic LDR and all of HDR spatial saliency models are introduced [61, 55].

2.6.1.1 LDR Saliency Models

Itti *et al.* proposed a saliency detection algorithm [58], which is the most widely used technique for LDR visual attention evaluation, the block diagram is displayed in Figure 2.28. This model predicts bottom-up attention from intensity, colour and orientation features in the Gaussian pyramid which are driven by HVS, then, the features are computed by centre-surround operations. The average of the conspicuous maps of each channel is the final saliency map. This model can be extended to the HDR image. However, it fails to detect the salient areas of the median range.

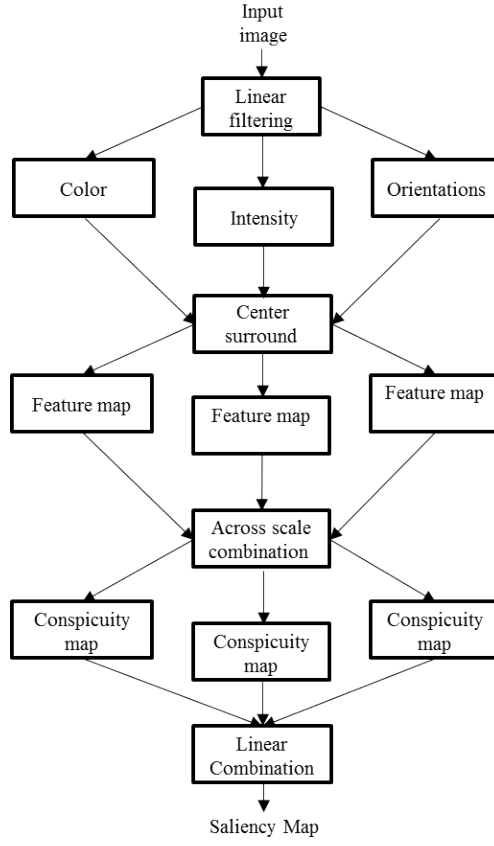


Figure 2.28: Architecture of the Itti *et al.* model

The colour channels are obtained by four broadly-tuned colour channels:

$$\begin{aligned}
 R &= r - (g + b)/2, \\
 G &= g - (r + b)/2, \\
 B &= b - (r + g)/2, \\
 Y &= (r + g)/2 - |r - g|/2 - b.
 \end{aligned} \tag{2.78}$$

$$\begin{aligned}
 RG(c, s) &= |(R(c) - G(c)) \ominus (G(s) - R(s))|, \\
 BY(c, s) &= |(B(c) - Y(c)) \ominus (Y(s) - B(s))|.
 \end{aligned} \tag{2.79}$$

where c is the pixel at the center scale, s is the pixel at the surrounding scale.

The cross-scale difference between two maps, defined as the ' \ominus ', is com-

puted by the interpolation to the greater scale and pixelwise subtraction.

$$\begin{aligned}
C'(c, s) &= |(RG(c) + BY(c)) \ominus (RG(s) + BY(s))|, \\
I'(c, s) &= |I(c) \ominus I(s)|, \\
O'(c, s, \theta) &= |O(c, \theta) \ominus O(s, \theta)|.
\end{aligned} \tag{2.80}$$

2.6.1.2 HDR Saliency Models

- Contrast Feature model:

The intensity channel in Itti *et al.* 's model has been replaced by the contrast features (CF) [17], because the biological sensors are more sensitive to the contrasts rather than absolute differences. Also, the orientation features are replaced by a closer HVS driven feature which is homogeneous to contrast. By following equations:

$$I''(c, s) = \frac{|I(c) \ominus I(s)|}{I(s)}, \tag{2.81}$$

$$O''(c, s, \theta) = \frac{O(c, \theta)}{I(s)}. \tag{2.82}$$

- CRF multi exposure model:

Gao *et al.* [45, 46] proposed an HDR saliency map prediction method by combining multiple saliency maps of Itti *et al.* 's model of LDR image which are virtually exposed from HDR input image. Then, these LDR saliency maps are combined by a weighting functions to generated the final saliency map of the given HDR input image. However, the camera response curve is required to virtual expose HDR image for different exposures.

The average logarithmic luminance is computed as:

$$L_{av} = \exp \left(\frac{1}{N} \sum_{x,y} \ln (L(x, y)) \right), \tag{2.83}$$

where N is total number of pixels, $L(x, y)$ is luminance values.

The L_{av} is then as a normalization factor of exposure which converts the logarithmic of average luminance to zero as followed:

$$L'(x, y) = \frac{L(x, y)}{L_{av}}. \quad (2.84)$$

Finally, the virtual photograph is mapped by:

$$L_d(x, y) = g^{-1}(\ln(L'(x, y) \Delta t)), \quad (2.85)$$

where g^{-1} is the inverse function of the response function which maps natural logarithm of exposure to the display pixel values L_d by exposure time Δt .

The response curve is known in the Gao *et al.*'s paper. Once virtual photographs are created, the saliency maps can be computed. The Itti *et al.*'s method is adopted without change.

The combination of saliency maps is solved as a weighting problem, it was updated in [46]. Inspired from [106], the estimation of the brightness of HDR image is computed by:

$$f = \frac{2 \log L_{av} - \log L_{min} - \log L_{max}}{\log L_{max} - \log L_{min}}. \quad (2.86)$$

The proper exposed luminance can be estimated by:

$$\bar{L} = a \frac{f + 1}{2}, \quad (2.87)$$

where a is 210 by approximation according to linear regression within [202, 219] as 95% confidence interval.

Then the corresponding exposure time for properly exposed luminance

can be calculated as:

$$\Delta \bar{t} = \Delta t * \frac{\bar{L}}{L}, \quad (2.88)$$

The weights for each virtual photographs are:

$$w(\Delta t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(\log_2 \frac{\Delta t}{\Delta \bar{t}})^2\right) \quad (2.89)$$

For example: for $\Delta t = [1/15, 1/8, 1/4, 1/2, 1, 2, 4, 8, 15]$, the corresponding $w = [0.005, 0.036, 0.147, 0.301, 0.308, 0.158, 0.040, 0.005, 0.0004]$.

Because of $\Delta \bar{t} = 0.72$, the closest 4th and 5th of w received larger weight than others.

- HDR features of Itti *et al.* model:

Dong *et al.* [30] proposed HVS based saliency detection method for HDR content (video). The final saliency map of this model is the fusion of spatial and temporal saliency maps. The spatial saliency is predicted on Itti *et al.* 's model with HDR input content pre-processing. The colour channels and luminance channel are transformed to Colour Appearance Model (CAM), and in the units of Just Noticeable Difference (JND) followed by CSF [21] filtering [85] respectively. The optical flow is adopted as motion descriptor to compute the temporal saliency map. The RGB colour space is converted to XYZ space, and then to LMS cone space:

The following is the Hunt-Pointer-Estevéz transform equations:

$$\begin{aligned} L' &= \frac{L^{n_c}}{L^{n_c} + L_a^{n_c}}, \\ M' &= \frac{M^{n_c}}{M^{n_c} + L_a^{n_c}}, \\ S' &= \frac{S^{n_c}}{S^{n_c} + L_a^{n_c}}, \end{aligned} \quad (2.90)$$

where L_a is the absolute level of adaptation luminance in cd/m^2 , and and the n_c is 0.57.

Two opponent-colour signals are derived as equations:

$$\begin{aligned} a &= \frac{1}{11}(11L' - 12M' + S'), \\ b &= \frac{1}{9}(L' + M' - S'). \end{aligned} \quad (2.91)$$

The input luminance is transformed into units of Just Noticeable Difference (JND) by equation:

$$luma(L_a) = \begin{cases} 769.18 \cdot L_a, & L_a < L_1 \\ 449.12 \cdot L_a^{0.17} - 232.25, & L_1 \leq L_a < L_2 \\ 181.7 \cdot \ln(L_a) - 90.16, & L_a \geq L_2 \end{cases}$$

$$L_1 = 0.061843 \quad cd/m^2, \quad L_2 = 164.1 \quad cd/m^2. \quad (2.92)$$

Dong suggested the multi CSF sensitivities with different

$L_a = (0.0001, 0.01, 0.1, 1, 10, 100, 1000)$, and they were applied to each Fourier transformed frequency domain.

Less salient areas are detected by Dong *et al.*'s model than Gao *et al.*'s, despite the fact that the CSF filtering step boost visible frequency, the CSF required specific tuning for particular application [5, 88].

These HDR spatial saliency models are evaluated in the Chapter 5 and the proposed HDR spatial saliency model has a over all better performance and less computational complexities.

2.6.2 Temporal Saliency Models

There are two types of temporal saliency model:

1. The bottom-up models use the motion information to predict the saliency of moving stimuli.
2. Other models aim to discriminate the event aspects of a task by machine learning methods mostly. They are usually top-down approaches, which are application-driven instead of stimulus-driven. This thesis emphasises on the

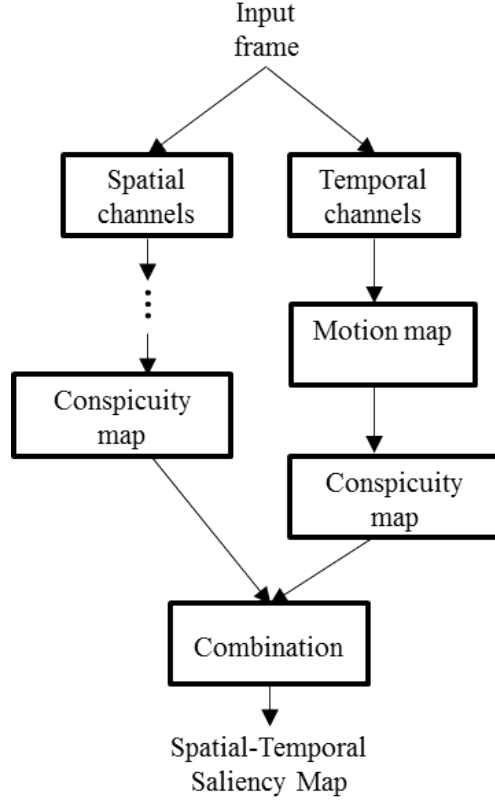


Figure 2.29: Architecture of the Spatial-Temporal model

first bottom-up approaches because they are stimulus-driven. Literature suggests for video [80] temporal saliency is the dominant contributor. The most of spatial-temporal saliency models use the following term or a variant to combine the spatial and temporal saliency map:

$$S = a \cdot S(s) + b \cdot S(t) + c \cdot S(s) \cdot S(t), \quad (2.93)$$

where a , b and c are weighting factors of spatial, temporal and spatial-temporal cross channel saliency maps, see Figure 2.29. The existing temporal saliency detection models are mostly based on one of the followings [72, 136, 105, 78, 10]: optical flow, frame differences and motion vector. The implementation detail varies and the performances are close to each others. [143, 71, 144, 79, 8, 93, 138]

2.6.3 Visual Saliency-Guided Compression

Lossy HDR image compressions JPEG-HDR [131], JPEG2000-HDR [135], JPEG-Xt [3] and Mapping-free HDR image compression [38] have been proposed. In the first three methods, the dynamic range is compressed to meet the input requirements of conventional compression codecs. In [38], the HDR format data presentation is unchanged and fed into conventional compression codecs directly. However, these HDR image compression approaches fail to compress non-salient region differently than the salient region, when HVS perceives them differently [107] especially for HDR content. The works of [101, 56, 49, 50] suggested LDR content can be compressed by visual saliency *i.e.* Region of Interest (ROI) in JPEG2000 [16, 77]. By extending this concept to HDR content, the HDR visual saliency-guided compression can be realized. However, the foveation methods [128, 56] produce blurring artifact, therefore the smoothing operation for the non-salient regions is ignored in proposed method.

2.6.4 Saliency Evaluation Metrics

The selection of saliency evaluation metrics are based on the findings of [70], which are Pearson Correlation Coefficient (PCC), Kullback-Leibler Divergence (KLD) Receiver and Operating Characteristic (ROC) Analysis.

2.6.4.1 Pearson Correlation Coefficient (CC)

The Pearson Correlation Coefficient can be computed as:

$$CC(A, B) = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{(\sum_m \sum_n (A_{mn} - \bar{A})^2)(\sum_m \sum_n (B_{mn} - \bar{B})^2)}}, \quad (2.94)$$

2.6.4.2 KLD

In Dong *et al.* 's paper, the KLD is computed by:

$$KL(A|B) = \sum_x A(x) \log \left(\frac{A(x)}{B(x)} \right), \quad (2.95)$$

where $A(x)$ and $B(x)$ are probability density functions (implemented by histogram) deduced from the prediction map and the ground truth map. This will give a positive result, and in the reference of Dong *et al.* 's paper [70], the equation was:

$$KL(A|B) = \sum_x A(x) \log \left(\frac{B(x)}{A(x)} \right), \quad (2.96)$$

and this will gives a negative result.

For a proper use of KLD which is described in [70], the KLD of two saliency map should be computed as:

$$\overline{KL(A|B) + KL(B|A)} = \overline{\sum_x A(x) \log \left(\frac{B(x)}{A(x)} \right) + \sum_x B(x) \log \left(\frac{A(x)}{B(x)} \right)}, \quad (2.97)$$

2.6.4.3 ROC

The ROC for continuous saliency maps can be computed as [70]:

$$\begin{aligned} TPr &= \frac{TP_i}{TP + FN}, \\ FPr &= \frac{FP_i}{FP + TN}, \end{aligned} \quad (2.98)$$

where i is for each threshold in saliency map from 0 to 100%.

2.6.5 Discussion

The LDR saliency model can be extended to HDR contents by adopting more discriminative features, *i.e.* CSF, CF, CRF and multi-exposures. However, the computational complexity increased. This is addressed in Chapter 5. Since the HDR media records the greater luminance range and wider colour gamut

than LDR media, the HDR contents are closer to the actual scene observed by HVS. Thus, saliency detection of HDR content represents more meaningful attention regions of the natural scene. The HDR applications could benefit from the saliency analysis *i.e.* compression, watermarking, TMO.

2.7 Evaluation Tools and Data-sets

In this section, the evaluation encoding tools, quality metrics and data sets are introduced. The thumbnails of data set are presented in the Appendix A.

2.7.1 Encoding Tools

The encoding tools used in this thesis are:

- JPEG–libJPEG ²
- JPEG2000–OpenJPEG ³
- JPEG–XT–JPEG XT ⁴
- Motion–JPEG and Motion–JPEG2000–Matlab R2016a
- HEVC–HM 16.0 Range Extension profile ⁵

2.7.2 Image dataset

The experiments are carried out on an image test-set containing 25 RGBE ⁶ [130], 9 EXR ⁷ [75] and 20 PFM ⁸ images [66]. They are widely used in many related types of research.

Table 2.4: Test-set Informations

²<https://github.com/thorfdbg/libjpeg>

³<https://github.com/uclouvain/openjpeg>

⁴<https://jpeg.org/jpegxt/software.html>

⁵<https://hevc.hhi.fraunhofer.de/rext>

⁶<http://www.anyhere.com/gward/hdrenc/pages/originals.html>

⁷<http://www.openexr.com/downloads>

⁸<http://mmspg.ep.ch/jpegxt-hdr>

| Image Name | Resolution | Size KB | Minimum Lumiance | Maximum Lumiance |
|-------------------------|------------|------------|------------------------|---------------------|
| Apartment.hdr | 1536x2048 | 9849.658 | 0 | 20.71835 |
| AtriumNight.hdr | 1016x760 | 2532.021 | 9.66×10^{-08} | 42 |
| Bottles.Small.hdr | 688x912 | 2451.073 | 0.000153 | 10.22267 |
| CS_Warwick.hdr | 512x1024 | 2048.074 | 0 | 6.894037 |
| Desk.hdr | 874x644 | 1874.974 | 0 | 87.5727 |
| Display1000.hdr | 1536x2048 | 10259.86 | 4.60E-07 | 26.22797 |
| Montreal.hdr | 1536x2048 | 9678.911 | 0.000937 | 40.75 |
| MtTamWest.hdr | 732x1214 | 2927.979 | 0.000834 | 9.656762 |
| Oxford_Church.hdr | 1200x1013 | 4748.513 | 0.000114 | 5.132082 |
| Spheron3.hdr | 1074x2149 | 5911.72 | 0 | 63851.82 |
| SpheronNapaValley.hdr | 2129x3025 | 18983.31 | 0.000364 | 83.2004 |
| SpheronNice.hdr | 1165x2981 | 10517.79 | 0.00107 | 416.2936 |
| SpheronPriceWestern.hdr | 1280x3272 | 13763.46 | 0.003305 | 17.64295 |
| SpheronSiggraph2001.hdr | 1289x1329 | 5235.805 | 0.001191 | 122 |
| StillLife.hdr | 846x1240 | 3301.952 | 0 | 8203.174 |
| Tree.hdr | 906x928 | 2878.355 | 0 | 10.03264 |
| bigFogMap.hdr | 1130x751 | 2681.899 | 0.003593 | 47.8417 |
| dani_belgium.hdr | 769x1025 | 2541.194 | 6.27×10^{-05} | 45.97645 |
| dani_cathedral.hdr | 1023x767 | 2610.869 | 0.000168 | 10.8125 |
| dani_synagogue.hdr | 769x1025 | 2449.326 | 0.016986 | 6.517875 |
| grand_canyon.hdr | 813x1249 | 3403.481 | 0.001263 | 0.322414 |
| memorial.hdr | 768x512 | 1309.914 | 0.001823 | 621.816 |
| nave.hdr | 480x720 | 1124.115 | 0 | 27185.43 |
| rosette.hdr | 480x720 | 1175.489 | 0 | 295.4368 |
| sundial.hdr | 1312x2000 | 8151.174 | 0.0004 | 0.195724 |
| CrissyField.exr | 810x1218 | 1274.042 | 0.001434 | 1.74248 |
| Flowers.exr | 734x784 | 740.315 | 0.015625 | 4.250468 |
| Garden.exr | 493x874 | 389.6934 | 0.004093 | 10.21094 |
| GoldenGate.exr | 860x1262 | 3607.04 | 0.000991 | 292.2596 |
| MtTamNorth.exr | 796x1198 | 1389.152 | 0.000298 | 7.873623 |
| Ocean.exr | 876x1255 | 4524.591 | 0.000218 | 1599.828 |
| Spirals.exr | 1040x1040 | 5238.721 | 0 | 5.912741 |
| StarField.exr | 1000x1000 | 1990.455 | 6.15×10^{-05} | 1272.261 |

| | | | | |
|---------------------------------|----------|----------|----------|----------|
| memorial.exr | 768x512 | 1110.644 | 0.000651 | 222.9744 |
| 507.pfm | 1080x944 | 11947.51 | 0.176359 | 4000.063 |
| BloomingGorse2.pfm | 1080x944 | 11947.51 | 2.847718 | 4000.063 |
| CanadianFalls.pfm | 1080x944 | 11947.51 | 4.542709 | 3269.765 |
| DevilsBathtub.pfm | 1080x944 | 11947.51 | 0.217655 | 4000.063 |
| HancockKitchenInside.pfm | 1080x944 | 11947.51 | 0.163662 | 3264.201 |
| LabTypewriter.pfm | 1080x944 | 11947.51 | 0.1713 | 3800.12 |
| LasVegasStore.pfm | 1080x944 | 11947.51 | 0.236267 | 4000.063 |
| McKeesPub.pfm | 1080x944 | 11947.51 | 0.164744 | 4000.063 |
| MtRushmore2.pfm | 1080x944 | 11947.51 | 0.163662 | 3609.217 |
| WillyDesk.pfm | 1080x944 | 11947.51 | 0.16628 | 4000.063 |
| dragon_3.pfm | 1080x944 | 11947.51 | 0.163662 | 4000.063 |
| set18.pfm | 1080x944 | 11947.51 | 0.163662 | 4000.063 |
| set22.pfm | 1080x944 | 11947.51 | 0.193852 | 3811.63 |
| set23.pfm | 1080x944 | 11947.51 | 0.343747 | 4000.063 |
| set24.pfm | 1080x944 | 11947.51 | 0.163662 | 3838.076 |
| set31.pfm | 1080x944 | 11947.51 | 0.198458 | 4000.063 |
| set33.pfm | 1080x944 | 11947.51 | 0.163662 | 4000.063 |
| set70.pfm | 1080x944 | 11947.51 | 0.169589 | 4000.063 |
| showgirl.pfm | 1080x944 | 11947.51 | 0.163662 | 3897.013 |
| sintel_2.pfm | 872x944 | 9646.52 | 0.603941 | 985.7157 |

2.7.3 Video Data-set

The video test-set ⁹ is provided by [42], each sequence of a scene is selected.

2.7.4 Eye Tracker Fixations

To validate the proposed HDR saliency model, the eye fixation dataset is used.

The ground truth comes from eye tracker dataset ¹⁰ [97].

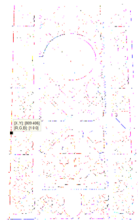
Only no. 14, 15, 17, 20, 22, 24, 33 and 39 are selected as evaluation dataset. because of other images have unreliable pixels. The example is given in Figure 2.30

⁹<https://hdr-2014.hdm-stuttgart.de>

¹⁰<http://mmspg.epfl.ch/hdr-eye>

Table 2.5: Video Test-set Infomations

| Sequence Name | Frames | Size MB | Minimum Lumiance | Maximum Lumiance |
|-----------------------|--------|---------|-------------------------|------------------|
| beerfest_lightshow_01 | 123 | 640 | 0 | 161.782 |
| bistro_01 | 151 | 1140 | -0.002 | 541.962 |
| carousel_fireworks_02 | 439 | 3740 | -2.75×10^{-05} | 439.849 |
| cars_longshot | 820 | 3870 | 0 | 458.621 |
| fireplace_01 | 491 | 3950 | -0.0067 | 0.6626 |
| fishing_longshot | 834 | 5660 | 0 | 10.548 |
| poker_fullshot | 600 | 3690 | -5.51×10^{-04} | 489.168 |
| showgirl_01 | 776 | 4380 | -0.003 | 459.502 |
| smith_hammering | 1102 | 7360 | -0.005 | 143.729 |



(a)



(b)

Figure 2.30: (a) Over saturated pixels; (b) zoomed-in tone-mapped version of the example

Chapter 3

DWT with TMO built-in HDR Visual Content Compression

In this Chapter, the relationship of DWT lifting steps and TMO are explored, and a novel approach to compress HDR image by JPEG2000 codec while backwards-compatible to LDR is proposed. This approach exploits the reversibility of TMO and scalability of DWT [121].

3.1 Motivation

The JPEG-HDR and JPEG2000-HDR are the two most widely accepted lossy HDR image compression standards, and the JPEG2000-HDR is superior in terms of compression performance due to the JPEG2000 codec. However, it is not compatible to the LDR since there is no tone-mapped LDR version encoded in the bitstream. In existing methods, the TMOs are performed before the decorrelation transforms *i.e.* DCT and DWT, it is because that the TMOs are not supposed to process the negative values *i.e.* negative coefficients of transforms. Hence, the applications working on the transformed frequency domain are compromised, *i.e.* watermarking and saliency estimation [98, 99, 9].

The possibility of combining the DWT and TMO is addressed by reviewing the decorrelation transforms, the TMOs and their corresponding ITMOs. In this chapter, an HDR compression scheme integrating TMO into DWT is proposed.

The requirements of propose HDR compression scheme are:

- Tone-mapping operation built-in to provide LDR compatibility.
- Each DWT decomposition level is processed in HDR values in order to preserve HDR characteristics of DWT decomposition.

To satisfy the above requirements, the input of DWT transform must be HDR. Moreover, considering that the TMOs are not supposed to map DWT coefficients. Therefore, the TMO must be applied before the end of DWT. To address this problem, the DWT with TMO built-in HDR compression is proposed.

3.1.1 DWT lifting theory

In this section, the preliminary DWT lifting theory is introduced, and the integration of TMO is demonstrated.

The DWT is usually implemented by filter bank and lifting scheme. The latter can be implemented faster with reduced memory requirements and can be easily inverted.

3.1.1.1 1D lifting

The lifting scheme proposed in [119, 120], can be described as the following parts:

- Input signal: $x = \{x_i \dots x_{2i+1}\}$, where $i \in \mathbb{N}$
- Lazy wavelet transform or polyphase decomposition of x_{2i} into two sub-signals:

An approximation signal a_i formed by the even samples of x_{2i} .

A detail signal b_i formed by the odd samples of x_{2i+1} .

- Lifting steps:

Prediction P lifting step that predicts the detail signal samples using the approximation samples a'_i ,

$$b'_i = b_i - P(a'_i). \quad (3.1)$$

Update U lifting step that updates the approximation signal with the detail samples b'_i

$$a'_i = a_i - U(b'_i). \quad (3.2)$$

- Output signal:

the transform coefficients a' and b' .

In practice, scaling factors K_a , K_b should be applied to normalize the transform coefficients, to ensure the input and output energy similarity.

Haar wavelet

Haar wavelet is the simplest wavelet which is also known as the first order Daubechies wavelet [22]. The polyphase domain analysis of filter banks permits us to obtain Perfect Reconstruction (PR) conditions on the filters and naturally leads to the lifting scheme as a built-in PR decomposition [23].

For Haar wavelets we have:

$$\begin{aligned} h(z) &= 1 + z^{-1}, \\ g(z) &= -\frac{1}{2} + \frac{1}{2}(z^{-1}) \\ \tilde{h}(z) &= \frac{1}{2} + \frac{1}{2}(z^{-1}), \\ \tilde{g}(z) &= -1 + z^{-1}. \end{aligned} \quad (3.3)$$

where \tilde{h}, \tilde{g} , are FIR filters of the wavelet transform. The polyphase matrix P as:

$$P(z) = \begin{bmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix}. \quad (3.4)$$

and analysis:

$$\tilde{P}(1/z) = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -1 & 1 \end{bmatrix}. \quad (3.5)$$

Which leaves $\tilde{P}^\top(1/z) = \begin{bmatrix} \frac{1}{2} & -1 \\ \frac{1}{2} & 1 \end{bmatrix}$

The implementation of corresponding forward transform when input signal x is split to even and odd samples $[a_i, b_i]$:

$$\begin{aligned} a_i &= x_{2i}, \\ b_i &= x_{2i+1}, \end{aligned} \quad (3.6)$$

$$\begin{aligned} b'_i &= x_{2i+1} - x_{2i}, \\ a'_i &= x_{2i} + \frac{1}{2}b'_i, \end{aligned} \quad (3.7)$$

If b'_i is substituted with $x_{2i+1} - x_{2i}$, then will get: $a'_i = \frac{1}{2}(x_{2i} + x_{2i+1})$, correspondingly, b'_i can be represented by a'_i and x_{2i+1} :

$$\begin{aligned} b'_i &= x_{2i+1} - x_{2i}, \\ &= -x_{2i} - x_{2i+1} + 2x_{2i+1}, \\ &= -2\left(\frac{1}{2}(x_{2i} + x_{2i+1})\right) + 2x_{2i+1}, \\ &= -2a'_i + 2x_{2i+1}, \end{aligned} \quad (3.8)$$

3.1.2 Extending 1D Haar lifting to 2D Haar lifting

The following equations are described by partitioning image into 2 by 2 non-overlap blocks and each pixel in each block represented as a , b , c and d where $i, j \in \mathbb{N}$ denotes the 2D index. They are corresponding to equation 3.6 (from

left to right, top to bottom) and also can be seen as a matrix [119, 120, 23, 2].

$$\begin{bmatrix} a_{i,j} & b_{i,j} \\ c_{i,j} & d_{i,j} \end{bmatrix}, \text{ and resulting matrix is } \begin{bmatrix} a'_{i,j} & b'_{i,j} \\ c'_{i,j} & d'_{i,j} \end{bmatrix}.$$

The \tilde{h} of equation 3.3 and \tilde{g} can be seen as $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ and $\begin{bmatrix} -1 & 1 \end{bmatrix}$ respectively.

Then the 2D filtering operation of the forward transform can be represented as:

$$\begin{aligned} \tilde{h} \cdot \tilde{h}^\top &= \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} \end{bmatrix} \\ \tilde{g} \cdot \tilde{h}^\top &= \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ \tilde{h} \cdot \tilde{g}^\top &= \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \\ \tilde{g} \cdot \tilde{g}^\top &= \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{aligned} \tag{3.9}$$

The coefficients of the forward transform are:

$$a'_{i,j} = \frac{1}{4}(a_{i,j} + b_{i,j} + c_{i,j} + d_{i,j}), \tag{3.10}$$

$$b'_{i,j} = \frac{1}{2}(a_{i,j} - b_{i,j} + c_{i,j} - d_{i,j}), \tag{3.11}$$

$$c'_{i,j} = \frac{1}{2}(a_{i,j} + b_{i,j} - c_{i,j} - d_{i,j}), \tag{3.12}$$

$$d'_{i,j} = \frac{1}{2}(a_{i,j} - b_{i,j} - c_{i,j} + d_{i,j}). \tag{3.13}$$

If substituted each next equation by using the results of the previous equations, the lifting steps are:

$$a'_{i,j} = \frac{1}{4}(a_{i,j} + b_{i,j} + c_{i,j} + d_{i,j}), \quad (3.14)$$

$$b'_{i,j} = -2a'_{i,j} + b_{i,j} + d_{i,j}, \quad (3.15)$$

$$c'_{i,j} = -2a'_{i,j} + c_{i,j} + d_{i,j}, \quad (3.16)$$

$$d'_{i,j} = -4a'_{i,j} - 2b'_{i,j} - 2c'_{i,j} + 4d_{i,j}. \quad (3.17)$$

The inverse transform:

$$d_{i,j} = a'_{i,j} + \frac{1}{2}b'_{i,j} + \frac{1}{2}c'_{i,j} + \frac{1}{4}d'_{i,j}. \quad (3.18)$$

$$c_{i,j} = 2a'_{i,j} + c'_{i,j} - d_{i,j}. \quad (3.19)$$

$$b_{i,j} = 2a'_{i,j} + b'_{i,j} - d_{i,j}. \quad (3.20)$$

$$a_{i,j} = 4a'_{i,j} - b_{i,j} - c_{i,j} - d_{i,j}. \quad (3.21)$$

Above $a'_{i,j}$, $b'_{i,j}$, $c'_{i,j}$ and $d'_{i,j}$ were representing LL, HL, LH, HH subbands of transformed image, see Figure 3.1. The LL will be an approximation of the original image, and HL, LH and HH will be horizontal, vertical and diagonal edges of the original image.

3.2 The Proposed method

This section starts by introducing the proposed lifting method with Haar lifting; then it is extended to 3/5 and 5/3 lifting schemes. Other lifting schemes can be implemented in a similar way.

3.2.1 DWT-TMO lifting

The lazy wavelet transform S down-sampled input HDR image I into 4 a, b, c and d even size part, where a, b, c and $d \geq 0$, because they usually represents

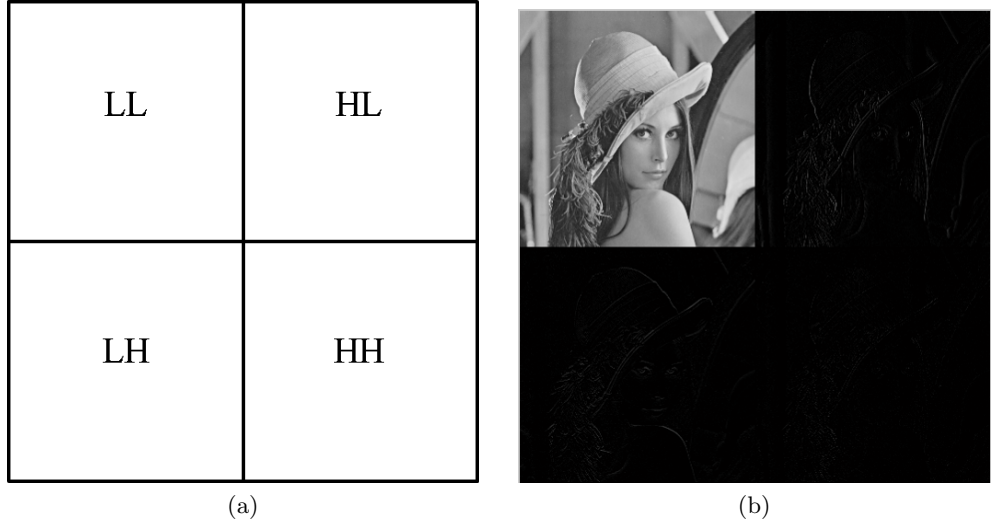


Figure 3.1: (a) An one level of DWT decomposition; (b) Lena example of one level of DWT decomposition.

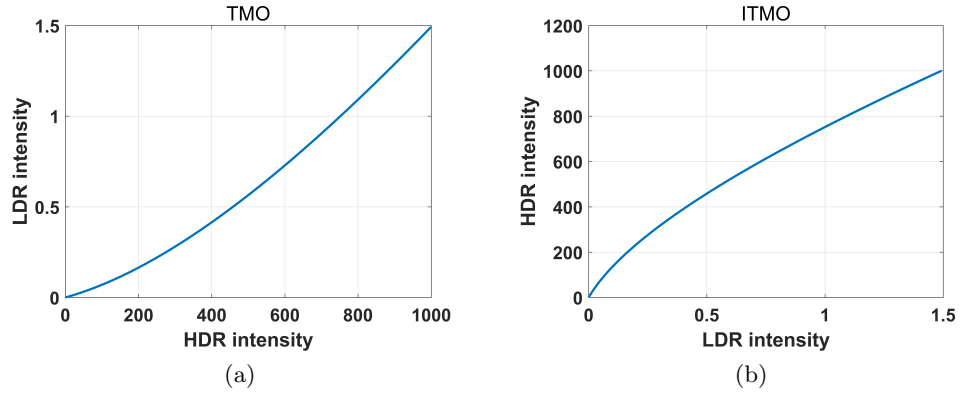


Figure 3.2: Reinhard Global TMO [108]
(a)TMO; (b)ITMO.

the intensity of the light. For most of global TMOs \mathcal{T} and their corresponding inverse functions – ITMOs \mathcal{T}' in Figure 3.2, then have:

$$D_l = \mathcal{T}(D) = S_D(\mathcal{T}(I)), \quad (3.22)$$

and

$$D = \mathcal{T}'(\mathcal{T}(D)) = S_D(I), \quad (3.23)$$

where D_l , represents tone-mapped of a , or b , or c or d , and D , represents HDR in the Figure 3.3 of a , or b , or c or d respectively. see Figure 3.4 and

3.5, the false colour map used is 'jet' in Matlab . This suggests that for each determined TMO curve, the results of a tone-mapped subband is not affected by the applying order of lazy wavelet transform and TMO.

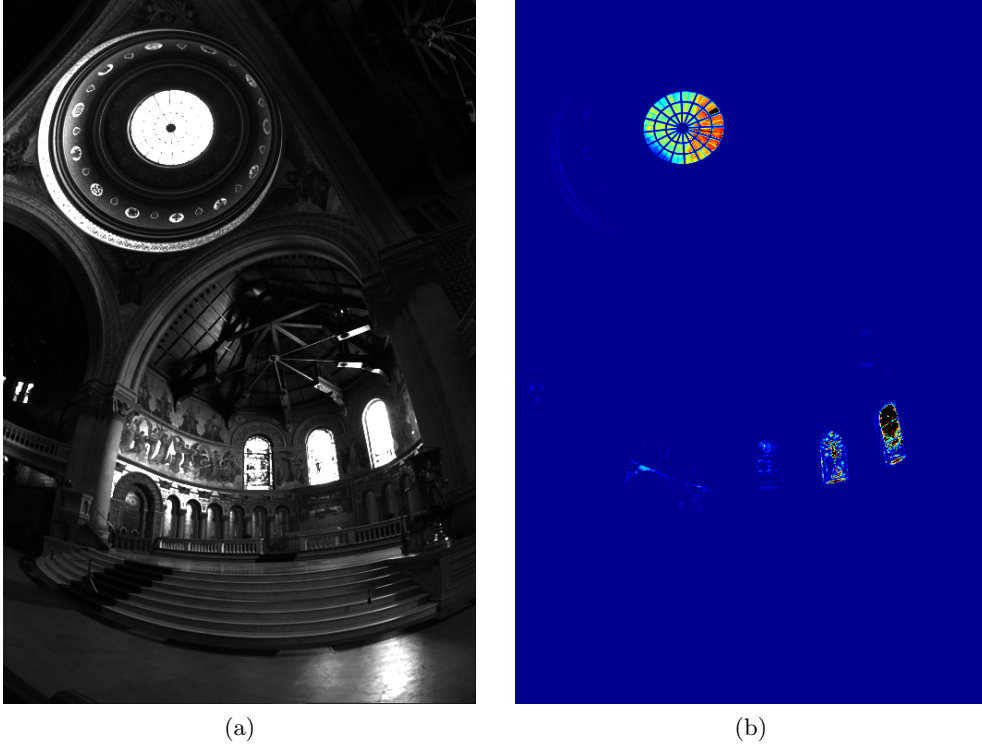


Figure 3.3: (a) HDR image ; (b) Jet colour map of HDR image .

Intuitively, there are two possible inputs for forward 2D Haar lifting: HDR and LDR.

HDR forward lifting:

$$a'_{i,j} = \frac{1}{4}(a_{i,j} + b_{i,j} + c_{i,j} + d_{i,j}), \quad (3.24)$$

$$b'_{i,j} = -2a'_{i,j} + b_{i,j} + d_{i,j}, \quad (3.25)$$

$$c'_{i,j} = -2a'_{i,j} + c_{i,j} + d_{i,j}, \quad (3.26)$$

$$d'_{i,j} = -4a'_{i,j} - 2b'_{i,j} - 2c'_{i,j} + 4d_{i,j}. \quad (3.27)$$

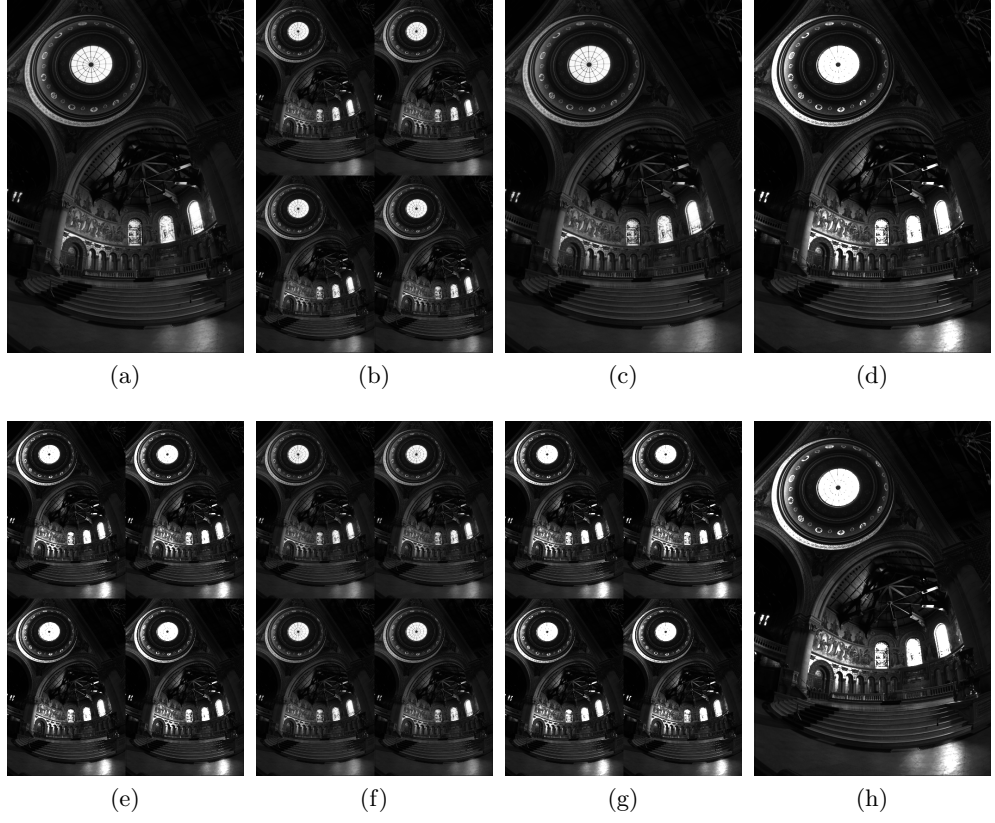


Figure 3.4: (a) Tone-mapped HDR image $\mathcal{T}(I)$; (b) Lazy wavelet transformed tone-mapped image $S(\mathcal{T}(I))$; (c) Inverse tone-mapped of (b) $\mathcal{T}'(S(\mathcal{T}(I)))$; (d) Inverse Lazy wavelet transformed of (c) $S'(\mathcal{T}'(S(\mathcal{T}(I))))$. (e) Lazy wavelet transformed HDR image $S(I)$; (f) Tone-mapped (b) $\mathcal{T}(S(I))$; (g) Inverse Lazy wavelet transformed of $S'(\mathcal{T}(S(I)))$; (h) Inverse tone-mapped of (g) $\mathcal{T}'(S'(\mathcal{T}(S(I))))$

LDR forward lifting:

$$a'_{li,j} = \frac{1}{4}(a_{li,j} + b_{li,j} + c_{li,j} + d_{li,j}), \quad (3.28)$$

$$b'_{li,j} = -2a'_{li,j} + b_{li,j} + d_{li,j}, \quad (3.29)$$

$$c'_{li,j} = -2a'_{li,j} + c_{li,j} + d_{li,j}, \quad (3.30)$$

$$d'_{li,j} = -4a'_{li,j} - 2b'_{li,j} - 2c'_{li,j} + 4d_{li,j}. \quad (3.31)$$

By replacing a_l , b_l , c_l and d_l with $\mathcal{T}(a)$, $\mathcal{T}(b)$, $\mathcal{T}(c)$ and $\mathcal{T}(d)$ from

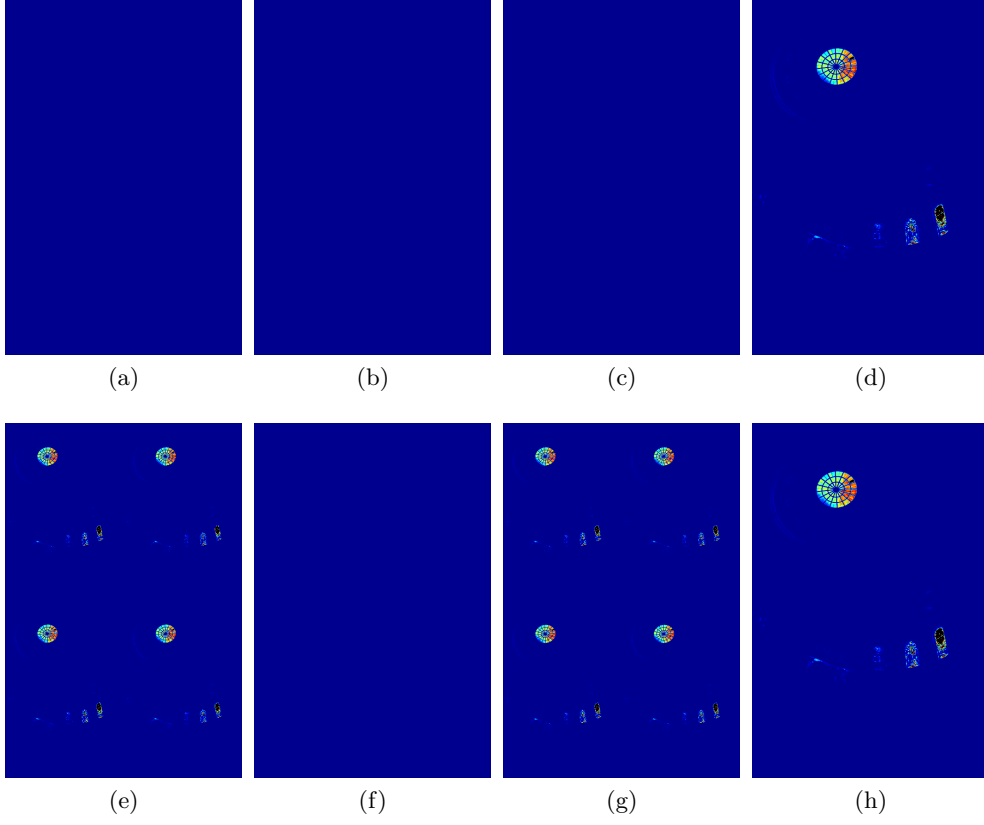


Figure 3.5: The corresponding false colour map of 3.4.

equation 3.22, where \mathcal{T} and \mathcal{T}' are TMO and ITMO, then have:

$$a'_{li,j} = \frac{1}{4}(\mathcal{T}(a_{i,j}) + \mathcal{T}(b_{i,j}) + \mathcal{T}(c_{i,j}) + \mathcal{T}(d_{i,j})), \quad (3.32)$$

$$b'_{li,j} = -2(a'_{li,j}) + \mathcal{T}(b_{i,j}) + \mathcal{T}(d_{i,j}), \quad (3.33)$$

$$c'_{li,j} = -2(a'_{li,j}) + \mathcal{T}(c_{i,j}) + \mathcal{T}(d_{i,j}), \quad (3.34)$$

$$d'_{li,j} = -4(a'_{li,j}) - 2(b'_{li,j}) - 2(c'_{li,j}) + 4\mathcal{T}(d_{i,j}). \quad (3.35)$$

From equation 3.32, 3.24 and 3.22 then have:

$$a'_{li,j} = \frac{1}{4} \mathcal{T}'(a'_{i,j}), \quad (3.36)$$

$$b'_{li,j} = -2\mathcal{T}(a'_{i,j}) + \mathcal{T}(b_{i,j}) + \mathcal{T}(d_{i,j}), \quad (3.37)$$

$$c'_{li,j} = -2\mathcal{T}(a'_{i,j}) + \mathcal{T}(c_{i,j}) + \mathcal{T}(d_{i,j}), \quad (3.38)$$

$$d'_{li,j} = -4\mathcal{T}(a'_{i,j}) - 2\mathcal{T}(b'_{li,j}) - 2\mathcal{T}(c'_{li,j}) + 4\mathcal{T}(d_{i,j}). \quad (3.39)$$

As above equations demonstrated, by keeping HDR LL subband and LDR HL , LH and HH subbands for each level of DWT decomposition, both HDR and LDR can be reconstructed in the same level. And the next level of DWT decomposition can be proceed with HDR LL subband. Since the coefficients of the low-pass filter of Haar are positive, the components of (a'_l) can be substituted by $\mathcal{T}(a')$, and the forward lifting of proposed method will always update first. However, 5/3 lifting predicts first, and it has negative coefficients of the low-pass filter, the examples are given in the following subsection.

3.2.1.1 Proposed Haar 2D DWT-TMO lifting

The proposed HDR image compression aims to apply TMO inside DWT lifting in order to satisfy the designing requirement. By analysing the equations of the forward Haar transform, the results suggest that the LL subband is consisted of summation of proportional down-sampled input signal, as illustrated in Figure 3.1 and 3.4. It means LL subband only contains the averaged HDR values, and they are positive values. Hence, this subband and 4 individual down-sampled input signals can be tone-mapped and inverse tone-mapped by a TMO \mathcal{T} and its corresponding inverse function \mathcal{T}' . For each next level decomposition, the input will be HDR instead of the tone-mapped version of the image. The following equations describe the proposed method on luminance channel only.

Proposed 2D Haar DWT-TMO forward lifting:

$$a'_{i,j} = \frac{1}{4}(a_{i,j} + b_{i,j} + c_{i,j} + d_{i,j}), \quad (3.40)$$

$$b'_{i,j} = -2\mathcal{T}(a'_{i,j}) + \mathcal{T}(b_{i,j}) + \mathcal{T}(d_{i,j}), \quad (3.41)$$

$$c'_{i,j} = -2\mathcal{T}(a'_{i,j}) + \mathcal{T}(c_{i,j}) + \mathcal{T}(d_{i,j}), \quad (3.42)$$

$$d'_{i,j} = -4\mathcal{T}(a'_{i,j}) - 2b'_{i,j} - 2c'_{i,j} + 4\mathcal{T}(d_{i,j}). \quad (3.43)$$

The inverse lifting steps are depending on whether HDR or LDR output is desired. Hence the proposed method contains LDR backwards-compatibility by integrating TMO in DWT lifting.

Proposed 2D Haar DWT-TMO inverse lifting: HDR

$$d_{i,j} = \mathcal{T}'(\mathcal{T}(a'_{i,j}) + \frac{1}{2}b'_{i,j} + \frac{1}{2}c'_{i,j} + \frac{1}{4}d'_{i,j}), \quad (3.44)$$

$$c_{i,j} = \mathcal{T}'(2\mathcal{T}(a'_{i,j}) + c'_{i,j} - \mathcal{T}(d_{i,j})), \quad (3.45)$$

$$b_{i,j} = \mathcal{T}'(2\mathcal{T}(a'_{i,j}) + b'_{i,j} - \mathcal{T}(d_{i,j})), \quad (3.46)$$

$$a_{i,j} = 4a'_{i,j} - b_{i,j} - c_{i,j} - d_{i,j}. \quad (3.47)$$

LDR

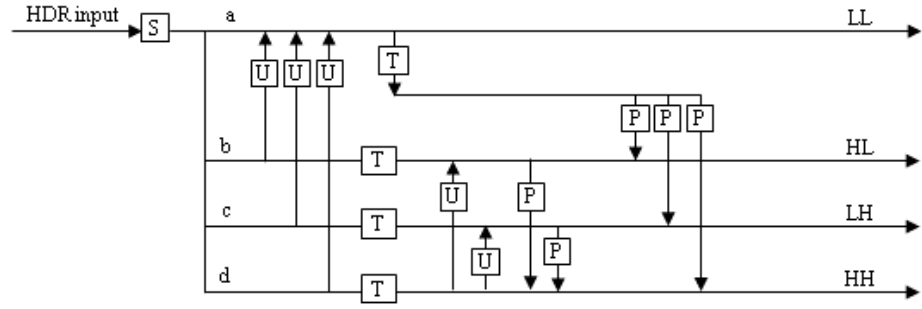
$$\mathcal{T}(d_{i,j}) = \mathcal{T}(a'_{i,j}) + \frac{1}{2}b'_{i,j} + \frac{1}{2}c'_{i,j} + \frac{1}{4}d'_{i,j}, \quad (3.48)$$

$$\mathcal{T}(c_{i,j}) = 2\mathcal{T}(a'_{i,j}) + c'_{i,j} - \mathcal{T}(d_{i,j}), \quad (3.49)$$

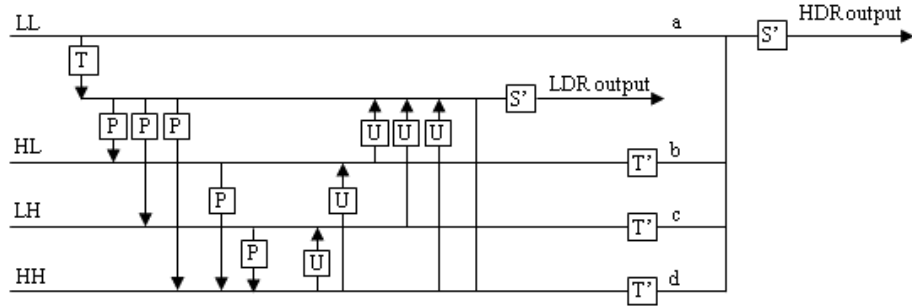
$$\mathcal{T}(b_{i,j}) = 2\mathcal{T}(a'_{i,j}) + b'_{i,j} - \mathcal{T}(d_{i,j}), \quad (3.50)$$

$$\mathcal{T}(a_{i,j}) = \mathcal{T}(4a'_{i,j}) - \mathcal{T}(b_{i,j}) - \mathcal{T}(c_{i,j}) - \mathcal{T}(d_{i,j}). \quad (3.51)$$

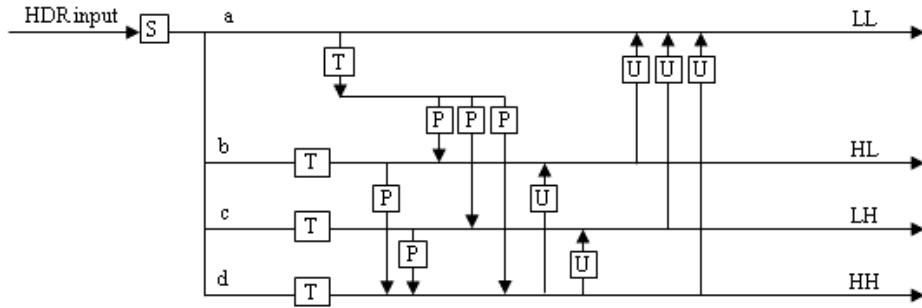
In Figure 3.6, the 2D lifting structures of the proposed methods are illustrated as update first example and prediction first example. Where the U , P , S , S' , T , T' represent, update, predict, LWT, ILWT, TMO, ITMO respectively.



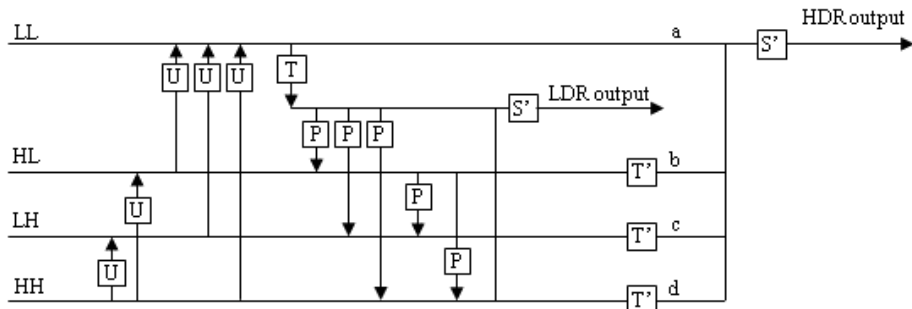
(a)



(b)



(c)



(d)

Figure 3.6: 2D lifting structure of the proposed method:
(a) Update first Forward lifting; (b) Update Inverse lifting; (c) Predict first Forward lifting; (d) Predict Inverse lifting.

3.2.1.2 Extending to 2D 3/5 and 2D 5/3 DWT-TMO lifting

2D 3/5 DWT-TMO lifting

The Haar lifting only uses 1 side of neighbour for predicting or updating. By implementing the predicting or updating operations of Haar based on geometrical symmetrical elements of 2 sides neighbours, the vanishing moment from Haar to 3/5 becomes 2 from 1. The lengths corresponding low-pass \tilde{h} and high-pass \tilde{g} filters became 3 and 5 respectively. The 2D 3/5 lifting matrix can be constructed from the 2D Haar lifting matrix:

$$\text{for block } \begin{bmatrix} d_{i-1,j-1} & c_{i,j-1} & d_{i,j-1} \\ b_{i-1,j} & \mathbf{a}'_{i,j} & b_{i,j} \\ d_{i-1,j} & c_{i,j} & d_{i,j} \end{bmatrix} \text{ Haar } \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \text{ and } 3/5 \begin{bmatrix} \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{16} \end{bmatrix}$$

then the forward lifting:

$$\begin{aligned} a'_{i,j} = & \frac{1}{4} \left(a_{i,j} + \frac{(b_{i-1,j} + b_{i,j})}{2} + \frac{(c_{i,j-1} + c_{i,j})}{2} \right. \\ & \left. + \frac{(d_{i-1,j-1} + d_{i-1,j} + d_{i,j-1} + d_{i,j})}{4} \right), \end{aligned} \quad (3.52)$$

$$\text{for block } \begin{bmatrix} c_{i,j-1} & d_{i,j-1} & c_{i+1,j-1} \\ \mathbf{a}'_{i,j} & b_{i,j} & a'_{i+1,j} \\ c_{i,j} & d_{i,j} & c_{i+1,j} \end{bmatrix} \text{ Haar } \begin{bmatrix} 0 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } 3/5 \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ -1 & 1 & -1 \\ 0 & \frac{1}{2} & 0 \end{bmatrix}$$

then the forward lifting:

$$b'_{i,j} = -2 \frac{(a'_{i,j} + a'_{i+1,j})}{2} + b_{i,j} + \frac{(d_{i,j} + d_{i,j-1})}{2}, \quad (3.53)$$

$$\text{for block } \begin{bmatrix} b_{i-1,j} & \mathbf{a}'_{i,j} & b_{i,j} \\ d_{i-1,j} & c_{i,j} & d_{i,j} \\ b_{i-1,j+1} & a'_{i,j+1} & b_{i+1,j+1} \end{bmatrix} \text{ Haar } \begin{bmatrix} 0 & -2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } 3/5 \begin{bmatrix} 0 & -1 & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & -1 & 0 \end{bmatrix}$$

then the forward lifting:

$$c'_{i,j} = -2 \frac{(a'_{i,j} + a'_{i,j+1})}{2} + c_{i,j} + \frac{(d_{i-1,j} + d_{i,j})}{2}, \quad (3.54)$$

$$\text{for block } \begin{bmatrix} a'_{i,j} & b'_{i,j} & a'_{i+1,j} \\ c'_{i,j} & d_{i,j} & c'_{i+1,j} \\ a'_{i,j+1} & b'_{i,j+1} & a'_{i+1,j+1} \end{bmatrix} \text{ Haar } \begin{bmatrix} -4 & -2 & 0 \\ -2 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } 3/5 \begin{bmatrix} -1 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

then the forward lifting:

$$d'_{i,j} = -4 \frac{(a'_{i,j} + a'_{i+1,j} + a'_{i,j+1} + a'_{i+1,j+1})}{4} - 2 \frac{(b'_{i,j} + b'_{i,j+1})}{2} - 2 \frac{(c'_{i,j} + c'_{i+1,j})}{2} + 4d_{i,j}. \quad (3.55)$$

The normal 2D 3/5 inverse lifting is given by:

$$d_{i,j} = \frac{d'_{i,j}}{4} + \frac{(a'_{i,j} + a'_{i+1,j} + a'_{i,j+1} + a'_{i+1,j+1})}{4} + \frac{(b'_{i,j} + b'_{i,j+1})}{4} + \frac{(c'_{i,j} + c'_{i+1,j})}{4}, \quad (3.56)$$

$$c_{i,j} = c'_{i,j} + (a'_{i,j} + a'_{i,j+1}) - \frac{(d_{i-1,j} + d_{i,j})}{2}, \quad (3.57)$$

$$b_{i,j} = b'_{i,j} + (a'_{i,j} + a'_{i+1,j}) - \frac{(d_{i,j} + d_{i,j-1})}{2}, \quad (3.58)$$

$$a_{i,j} = 4a'_{i,j} - \frac{(b_{i-1,j} + b_{i,j})}{2} - \frac{(c_{i,j-1} + c_{i,j})}{2} - \frac{(d_{i-1,j-1} + d_{i-1,j} + d_{i,j-1} + d_{i,j})}{4}, \quad (3.59)$$

Similarly, the LL subband a' and 4 individual down-sampled input signals a, b, c, d can be tone-mapped and inverse tone-mapped by a tone-mapping operator \mathcal{T} and its corresponding inverse function \mathcal{T} . For each next level decomposition, the input will be HDR instead of the tone-mapped version of the image.

Proposed 2D 3/5 DWT-TMO forward lifting:

$$\begin{aligned}
& \text{for block } \begin{bmatrix} d_{i-1,j-1} & c_{i,j-1} & d_{i,j-1} \\ b_{i-1,j} & \mathbf{a}_{i,j} & b_{i,j} \\ d_{i-1,j} & c_{i,j} & d_{i,j} \end{bmatrix} \\
& a'_{i,j} = \frac{1}{4} \left(a_{i,j} + \frac{(b_{i-1,j} + b_{i,j})}{2} + \frac{(c_{i,j-1} + c_{i,j})}{2} \right. \\
& \quad \left. + \frac{(d_{i-1,j-1} + d_{i-1,j} + d_{i,j-1} + d_{i,j})}{4} \right), \tag{3.60}
\end{aligned}$$

$$\begin{aligned}
& \text{for block } \begin{bmatrix} c_{i,j-1} & d_{i,j-1} & c_{i+1,j-1} \\ \mathbf{a}_{i,j} & b_{i,j} & a'_{i+1,j} \\ c_{i,j} & d_{i,j} & c_{i+1,j} \end{bmatrix} \\
& b'_{i,j} = -2 \frac{(\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j}))}{2} + \mathcal{T}(b_{i,j}) + \frac{(\mathcal{T}(d_{i,j}) + \mathcal{T}(d_{i,j-1}))}{2}, \tag{3.61}
\end{aligned}$$

$$\begin{aligned}
& \text{for block } \begin{bmatrix} b_{i-1,j} & \mathbf{a}'_{i,j} & b_{i,j} \\ d_{i-1,j} & c_{i,j} & d_{i,j} \\ b_{i-1,j+1} & a'_{i,j+1} & b_{i+1,j+1} \end{bmatrix} \\
& c'_{i,j} = -2 \frac{(\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i,j+1}))}{2} + \mathcal{T}(c_{i,j}) + \frac{(\mathcal{T}(d_{i-1,j}) + \mathcal{T}(d_{i,j}))}{2}, \tag{3.62}
\end{aligned}$$

$$\begin{aligned}
& \text{for block } \begin{bmatrix} \mathbf{a}'_{i,j} & b'_{i,j} & a'_{i+1,j} \\ c'_{i,j} & d_{i,j} & c'_{i+1,j} \\ a'_{i,j+1} & b'_{i,j+1} & a'_{i+1,j+1} \end{bmatrix} \\
& d'_{i,j} = -4 \frac{(\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j}) + \mathcal{T}(a'_{i,j+1}) + \mathcal{T}(a'_{i+1,j+1}))}{4} \\
& \quad - 2 \frac{(\mathcal{T}(b'_{i,j}) + \mathcal{T}(b'_{i,j+1}))}{2} - 2 \frac{(\mathcal{T}(c'_{i,j}) + \mathcal{T}(c'_{i+1,j}))}{2} \\
& \quad + 4 \mathcal{T}(d_{i,j}). \tag{3.63}
\end{aligned}$$

The inverse lifting steps are depending on whether HDR or LDR output is desired. Hence the proposed method contains LDR backwards-compatibility by integrating TMO in DWT lifting.

Proposed 2D 3/5 DWT-TMO inverse lifting:

HDR

$$d_{i,j} = \mathcal{T}'\left(\frac{d'_{i,j}}{4} + \frac{\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j}) + \mathcal{T}(a'_{i,j+1}) + \mathcal{T}(a'_{i+1,j+1})}{4} + \frac{b'_{i,j} + b'_{i,j+1}}{4} + \frac{(c'_{i,j} + c'_{i+1,j})}{4}\right), \quad (3.64)$$

$$c_{i,j} = \mathcal{T}'(c'_{i,j} + (\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j})) - \frac{(\mathcal{T}(d_{i-1,j}) + \mathcal{T}(d_{i,j}))}{2}), \quad (3.65)$$

$$b_{i,j} = \mathcal{T}'(b'_{i,j} + (\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j})) - \frac{(\mathcal{T}(d_{i,j}) + \mathcal{T}(d_{i,j-1}))}{2}), \quad (3.66)$$

$$a_{i,j} = 4a'_{i,j} - \frac{b_{i-1,j} + b_{i,j}}{2} - \frac{c_{i,j-1} + c_{i,j}}{2} - \frac{d_{i-1,j-1} + d_{i-1,j} + d_{i,j-1} + d_{i,j}}{4}, \quad (3.67)$$

LDR

$$\mathcal{T}(d_{i,j}) = \frac{d'_{i,j}}{4} + \frac{\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j}) + \mathcal{T}(a'_{i,j+1}) + \mathcal{T}(a'_{i+1,j+1})}{4} + \frac{b'_{i,j} + b'_{i,j+1}}{4} + \frac{(c'_{i,j} + c'_{i+1,j})}{4}, \quad (3.68)$$

$$\mathcal{T}(c_{i,j}) = c'_{i,j} + (\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j})) - \frac{(\mathcal{T}(d_{i-1,j}) + \mathcal{T}(d_{i,j}))}{2}, \quad (3.69)$$

$$\mathcal{T}(b_{i,j}) = b'_{i,j} + (\mathcal{T}(a'_{i,j}) + \mathcal{T}(a'_{i+1,j})) - \frac{(\mathcal{T}(d_{i,j}) + \mathcal{T}(d_{i,j-1}))}{2}, \quad (3.70)$$

$$\mathcal{T}(a_{i,j}) = 4\mathcal{T}(a'_{i,j}) - \frac{(\mathcal{T}(b_{i-1,j}) + \mathcal{T}(b_{i,j}))}{2} - \frac{(\mathcal{T}(c_{i,j-1}) + \mathcal{T}(c_{i,j}))}{2} - \frac{(\mathcal{T}(d_{i-1,j-1}) + \mathcal{T}(d_{i-1,j}) + \mathcal{T}(d_{i,j-1}) + \mathcal{T}(d_{i,j}))}{4}, \quad (3.71)$$

2D 5/3 DWT-TMO lifting

The low-pass and high-pass filters of 5/3 DWT are $\tilde{h} = \begin{bmatrix} -\frac{1}{8} & \frac{2}{8} & \frac{6}{8} & \frac{2}{8} & -\frac{1}{8} \end{bmatrix}$ and $\tilde{g} = \begin{bmatrix} -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix}$ respectively. Then the 2D filtering operation of the

forward transform can represent as:

$$\begin{aligned}
\tilde{h} \cdot \tilde{h}^\top &= \begin{bmatrix} \frac{1}{64} & -\frac{1}{32} & -\frac{3}{32} & -\frac{1}{32} & \frac{1}{64} \\ -\frac{1}{32} & \frac{1}{16} & \frac{3}{16} & \frac{1}{16} & -\frac{1}{32} \\ -\frac{3}{32} & \frac{3}{16} & \frac{9}{16} & \frac{3}{16} & -\frac{3}{32} \\ -\frac{1}{32} & \frac{1}{16} & \frac{3}{16} & \frac{1}{16} & -\frac{1}{32} \\ \frac{1}{64} & -\frac{1}{32} & -\frac{3}{32} & -\frac{1}{32} & \frac{1}{64} \end{bmatrix} \\
\tilde{g} \cdot \tilde{h}^\top &= \begin{bmatrix} \frac{1}{16} & -\frac{1}{8} & \frac{1}{16} \\ -\frac{1}{8} & \frac{1}{4} & -\frac{1}{8} \\ -\frac{3}{8} & \frac{3}{4} & -\frac{3}{8} \\ -\frac{1}{8} & \frac{1}{4} & -\frac{1}{8} \\ \frac{1}{16} & -\frac{1}{8} & \frac{1}{16} \end{bmatrix} \\
\tilde{h} \cdot \tilde{g}^\top &= \begin{bmatrix} \frac{1}{16} & -\frac{1}{8} & -\frac{3}{8} & -\frac{1}{8} & \frac{1}{16} \\ -\frac{1}{8} & \frac{1}{4} & \frac{3}{4} & \frac{1}{4} & -\frac{1}{8} \\ \frac{1}{16} & -\frac{1}{8} & -\frac{3}{8} & -\frac{1}{8} & \frac{1}{16} \end{bmatrix} \\
\tilde{g} \cdot \tilde{g}^\top &= \begin{bmatrix} \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix}
\end{aligned} \tag{3.72}$$

Normal 2D 5/3 DWT-TMO forward lifting (the 5/3 wavelets transform starts with prediction rather than update):

$$\begin{aligned}
&\text{for block } \begin{bmatrix} \mathbf{a}_{i,j} & b_{i,j} & a_{i+1,j} \\ c_{i,j} & d_{i,j} & c_{i+1,j} \\ a_{i,j+1} & b_{i+1,j+1} & a_{i+1,j+1} \end{bmatrix} \\
d'_{i,j} &= \frac{(a_{i,j} + a_{i+1,j} + a_{i,j+1} + a_{i+1,j+1})}{4} \\
&\quad - \frac{(b_{i,j} + b_{i,j+1})}{2} - \frac{(c_{i,j} + c_{i+1,j})}{2} + d_{i,j},
\end{aligned} \tag{3.73}$$

for block $\begin{bmatrix} b_{i-1,j} & \mathbf{a}_{i,j} & b_{i,j} \\ d'_{i-1,j} & c_{i,j} & d'_{i,j} \\ b_{i-1,j+1} & a_{i,j+1} & b_{i+1,j+1} \end{bmatrix}$ by interpret given block without sub-

script to: $\begin{bmatrix} a & b & a & b & a \\ c & d & c & d & c \\ a & b & a & b & a \end{bmatrix}$

$$\begin{aligned}
c' &= -\frac{1}{2}a - \frac{1}{2}b + \frac{1}{2}c + \frac{1}{2}d, \\
c' &= -\frac{1}{2}(a + b - c - d), \\
d' &= a - b - c + d, \\
d &= d' + c + b - a, \\
c' &= -\frac{1}{2}(a + b - c - d' - c - b + a), \\
c' &= c - a + \frac{1}{2}d',
\end{aligned} \tag{3.74}$$

by substituting d with d' without subscript, the c' is derived as above, then by sharing weight for the geometrical symmetrical corresponding element to the center:

$$c'_{i,j} = -\frac{(a_{i,j} + a_{i,j+1})}{2} + c_{i,j} + \frac{(d'_{i-1,j} + d'_{i,j})}{4}, \tag{3.75}$$

for block $\begin{bmatrix} c_{i,j-1} & d'_{i,j-1} & c_{i+1,j-1} \\ \mathbf{a}_{i,j} & b_{i,j} & a'_{i+1,j} \\ c_{i,j} & d'_{i,j} & c_{i+1,j} \end{bmatrix}$

by interpret given block without subscript to:

$$\begin{bmatrix} a & b & a \\ c & d & c \\ a & b & a \\ c & d & c \\ a & b & a \end{bmatrix}$$

$$\begin{aligned} b' &= -\frac{1}{2}a + \frac{1}{2}b - \frac{1}{2}c + \frac{1}{2}d, \\ b' &= -\frac{1}{2}(a - b + c - d), \\ d' &= a - b - c + d, \\ d &= d' + c + b - a, \\ b' &= -\frac{1}{2}(a - b + c - d' - c - b + a), \\ b' &= b - a + \frac{1}{2}d', \end{aligned} \tag{3.76}$$

by substituting d with d' without subscript, the b' is derived as above, then by sharing weight for the geometrical symmetrical corresponding element to the center:

$$b'_{i,j} = -\frac{(a_{i,j} + a_{i+1,j})}{2} + b_{i,j} + \frac{(d'_{i,j} + d'_{i,j-1})}{4}, \tag{3.77}$$

for block

$$\begin{bmatrix} d'_{i-1,j-1} & c'_{i,j-1} & d'_{i,j-1} \\ b'_{i-1,j} & \mathbf{a}_{i,j} & b'_{i,j} \\ d'_{i-1,j} & c'_{i,j} & d'_{i,j} \end{bmatrix}$$

by interpreting given block without subscript to:

$$\begin{bmatrix} a & b & a & b & a \\ c & d & c & d & c \\ a & b & a & b & a \\ c & d & c & d & c \\ a & b & a & b & a \end{bmatrix}$$

$$a' = \frac{1}{4}a + \frac{1}{4}b + \frac{1}{4}c + \frac{1}{4}d,$$

$$d = d' + c + b - a,$$

$$b = b' + a - \frac{1}{2}d',$$

$$c = c' + a - \frac{1}{2}d',$$

by replacing d with d' (3.78)

$$a' = \frac{1}{4}(2b + 2c + d'),$$

by replacing b, c with b', c'

$$a' = \frac{1}{4}(4a + 2b' + 2c' - d'),$$

$$a' = a + \frac{1}{2}b' + \frac{1}{2}c' - \frac{1}{4}d',$$

by substituting d, c, b with d', c', b' without subscript, the a' is derived as above. The vanishing moment from Haar to 5/3 becomes 2 from 1. By sharing weight for the geometrical symmetrical corresponding element to the center:

$$a'_{i,j} = a_{i,j} + \frac{(b'_{i-1,j} + b'_{i,j})}{4} + \frac{(c'_{i,j-1} + c'_{i,j})}{4} - \frac{(d'_{i-1,j-1} + d'_{i-1,j} + d'_{i,j-1} + d'_{i+1,j+1})}{8}, \quad (3.79)$$

Normal 2D 5/3 inverse lifting:

$$a_{i,j} = a'_{i,j} - \frac{(b'_{i-1,j} + b'_{i,j})}{4} - \frac{(c'_{i,j-1} + c'_{i,j})}{4} + \frac{(d'_{i-1,j-1} + d'_{i-1,j} + d'_{i,j-1} + d'_{i,j})}{8}, \quad (3.80)$$

$$b_{i,j} = \frac{(a_{i,j} + a_{i+1,j})}{2} + b'_{i,j} + \frac{(d'_{i,j} - d'_{i,j-1})}{4}, \quad (3.81)$$

$$c_{i,j} = \frac{(a_{i,j} + a_{i,j+1})}{2} + c'_{i,j} - \frac{(d'_{i-1,j} + d'_{i,j})}{4}, \quad (3.82)$$

$$d_{i,j} = -\frac{(a_{i,j} + a_{i+1,j} + a_{i,j+1} + a_{i+1,j+1})}{4} + \frac{(b_{i,j} + b_{i,j+1})}{2} + \frac{(c_{i,j} + c_{i+1,j})}{2} + d'_{i,j}, \quad (3.83)$$

The 5/3 wavelets transform starts with prediction rather than update. Hence, the LL subband here is computed by d', c', b', a' and 4 individual down-sampled input signals a, b, c, d . These down-sampled input signals can be tone-mapped and inverse tone-mapped by a tone-mapping operator \mathcal{T} and its' corresponding inverse function \mathcal{T} . For each next level decomposition, the input will be HDR instead of a tone-mapped version of the image.

Proposed 2D 5/3 DWT-TMO forward lifting:

$$\text{for block } \begin{bmatrix} \mathbf{a}_{i,j} & b_{i,j} & a_{i+1,j} \\ c_{i,j} & d_{i,j} & c_{i+1,j} \\ a_{i,j+1} & b_{i+1,j+1} & a_{i+1,j+1} \end{bmatrix}$$

$$d'_{i,j} = \frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i+1,j}) + \mathcal{T}(a_{i,j+1}) + \mathcal{T}(a_{i+1,j+1}))}{4} - \frac{(\mathcal{T}(b_{i,j}) + \mathcal{T}(b_{i,j+1}))}{2} - \frac{(\mathcal{T}(c_{i,j}) + \mathcal{T}(c_{i+1,j}))}{2} + \mathcal{T}(d_{i,j}), \quad (3.84)$$

$$\text{for block } \begin{bmatrix} b_{i-1,j} & \mathbf{a}_{i,j} & b_{i,j} \\ d'_{i-1,j} & c_{i,j} & d'_{i,j} \\ b_{i-1,j+1} & a'_{i,j+1} & b_{i+1,j+1} \end{bmatrix}$$

$$c'_{i,j} = -\frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i,j+1}))}{2} + \mathcal{T}(c_{i,j}) + \frac{(d'_{i-1,j} + d'_{i,j})}{4}, \quad (3.85)$$

$$\text{for block } \begin{bmatrix} c_{i,j-1} & d'_{i,j-1} & c_{i+1,j-1} \\ \mathbf{a}_{i,j} & b_{i,j} & a'_{i+1,j} \\ c_{i,j} & d'_{i,j} & c_{i+1,j} \end{bmatrix}$$

$$b'_{i,j} = -\frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i+1,j}))}{2} + \mathcal{T}(b_{i,j}) + \frac{(d'_{i,j} + d'_{i,j-1})}{4}, \quad (3.86)$$

$$\text{for block } \begin{bmatrix} d'_{i-1,j-1} & c'_{i,j-1} & d'_{i,j-1} \\ b'_{i-1,j} & \mathbf{a}_{i,j} & b'_{i,j} \\ d'_{i-1,j} & c'_{i,j} & d'_{i,j} \end{bmatrix}$$

$$a'_{i,j} = a_{i,j} + \frac{(b'_{i-1,j} + b'_{i,j})}{4} + \frac{(c'_{i,j-1} + c'_{i,j})}{4} - \frac{(d'_{i-1,j-1} + d'_{i-1,j} + d'_{i,j-1} + d'_{i,j})}{8}, \quad (3.87)$$

The inverse lifting steps are depending on whether HDR or LDR output is desired. Hence the proposed method contains LDR backwards-compatibility by integrating TMO in DWT lifting.

Proposed 2D 5/3 DWT-TMO inverse lifting:

HDR

$$a_{i,j} = a'_{i,j} - \frac{(b'_{i-1,j} + b'_{i,j})}{4} - \frac{(c'_{i,j-1} + c'_{i,j})}{4} + \frac{(d'_{i-1,j-1} + d'_{i-1,j} + d'_{i,j-1} + d'_{i,j})}{8}, \quad (3.88)$$

$$b_{i,j} = \mathcal{T}'\left(\frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i+1,j}))}{2}\right) + b'_{i,j} - \frac{(d'_{i,j} + d'_{i,j-1})}{4}, \quad (3.89)$$

$$c_{i,j} = \mathcal{T}'\left(\frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i,j+1}))}{2}\right) + c'_{i,j} - \frac{(d'_{i-1,j} + d'_{i,j})}{4}, \quad (3.90)$$

$$d_{i,j} = \mathcal{T}'\left(-\frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i+1,j}) + \mathcal{T}(a_{i,j+1}) + \mathcal{T}(a_{i+1,j+1}))}{4}\right) + \frac{(\mathcal{T}(b_{i,j}) + \mathcal{T}(b_{i,j+1}))}{2} + \frac{(\mathcal{T}(c_{i,j}) + \mathcal{T}(c_{i+1,j}))}{2} + d'_{i,j}, \quad (3.91)$$

LDR

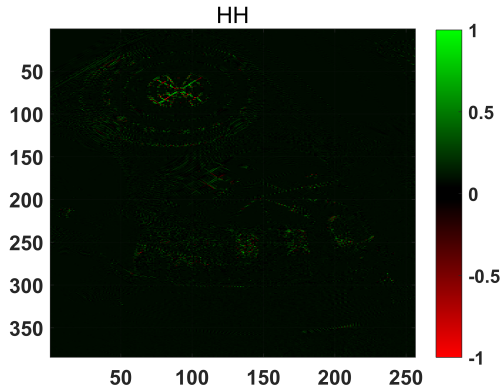
$$\mathcal{T}(a_{i,j}) = \mathcal{T}\left(a'_{i,j} - \frac{(b'_{i-1,j} + b'_{i,j})}{4} - \frac{(c'_{i,j-1} + c'_{i,j})}{4} + \frac{(d'_{i-1,j-1} + d'_{i-1,j} + d'_{i,j-1} + d'_{i,j})}{8}\right), \quad (3.92)$$

$$\mathcal{T}(b_{i,j}) = \frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i+1,j}))}{2} + b'_{i,j} - \frac{(d'_{i,j} + d'_{i,j-1})}{4}, \quad (3.93)$$

$$\mathcal{T}(c_{i,j}) = \frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i,j+1}))}{2} + c'_{i,j} - \frac{(d'_{i-1,j} + d'_{i,j})}{4}, \quad (3.94)$$

$$\mathcal{T}(d_{i,j}) = -\frac{(\mathcal{T}(a_{i,j}) + \mathcal{T}(a_{i+1,j}) + \mathcal{T}(a_{i,j+1}) + \mathcal{T}(a_{i+1,j+1}))}{4} + \frac{(\mathcal{T}(b_{i,j}) + \mathcal{T}(b_{i,j+1}))}{2} + \frac{(\mathcal{T}(c_{i,j}) + \mathcal{T}(c_{i+1,j}))}{2} + d'_{i,j}, \quad (3.95)$$

Compensation Procedures



(a)



(b)

Figure 3.7: (a) HH subband; (b) tone-mapped HH subband.

The 5/3 wavelets transform starts with prediction rather than update. Moreover, because of the negative coefficients of the low-pass filter, the filtered results could be negative. The proposed scheme of Haar and 3/5 could not be applied directly. The TMOs are not supposed to handle negative values, because the TMOs are an approximation of logarithmic functions without negative input. Therefore there is no real (\mathbb{R}) corresponding result for the

ITMO, as shown by an example in Figure 3.7. The proposed method uses positive values to proceed the next level of 2D DWT lifting and record negative values as an additional layer. The error will be introduced after quantization, and the results of inverse lifting could contain negative values as well. All negative values are ignored when applying TMO and ITMO. Other liftings which begin with prediction step first can be implemented in similarly as 5/3, such as CohenDaubechiesFeauveau (CDF) 9/7¹ applied at JPEG2000 lossy mode.

3.3 Experimental Results

The experimental setups and results are presented below. The full Rate Distortion (RD) results are demonstrated in the Appendix.

3.3.1 HDR test image information

The test-sets information was introduced in Section 2.7, the experiments in this chapter are based on all RGBE images and EXR images.

3.3.2 Experimental setups

Results in this section are produced by 3-levels 2D lifting on luminance channel.

The traditional method for comparison is tone-mapping first and do 3-levels of DWT-lifting, which is similar to JPEG2000-HDR and JPEG-HDR.

The results of [52] demonstrated that MSE and HDR-VDP are two most reliable metrics which are close to the subjective ground truth. As result, the MSE and HDR-VDP metrics are used in this thesis as HDR image quality assessments to evaluate the compression distortions.

The HDR-VDP metrics settings: no masking, luminance channel only, the viewing distance is 80 centimeters, the screen resolution is 1920 by 1200, the

¹M. Unser and T. Blu. "Mathematical Properties of the JPEG2000 Wavelet Filters." IEEE Trans. on Image Proc., vol. 12, no. 9, Sep. 2003

screen size is 24 inches, the visual cycles per degree is 50.

3.3.2.1 TMO

The proposed method has the flexibility that allows any TMO to be applied as long as it has the corresponding inverse TMO. The Reinhard's [108] photographic operator global approach is adopted because it obtained a balance between complexity and performance compared with other TMOs and has automatic parameter estimation mentioned in Chapter 2. Another important reason is it provides completely invertible inverse tone-mapping with minimal storage cost.

3.3.2.2 Quantization method

Different to the conventional quantization methods which are integer dividing and rounding, the proposed quantization sorts coefficients into different integer bins.

$$Q_s = \lfloor \frac{C_{max} - C_{min}}{Q_f} \rfloor, \quad (3.96)$$

where C is the input DWT coefficient, Q_s and Q_f are quantization step and quantization factor respectively. Once Q_s is determined, the quantized value Q is given by:

$$Q = \lfloor \frac{I}{Q_s} \rfloor, \quad (3.97)$$

where I is the coefficient value.

Dequantization procedure is the multiplied result of the order number of a bin and half of the bin size (mid value).

$$I' = QQ_s, \quad (3.98)$$

The maximum quantization error is limited by half of the bin size. Quantization factors (the numbers of bins) for RGBE files are as 32, 64, 128, 256, 512, 1024, 2000, 4000, 8000, 16000, 50000. The same quantization steps were

repeated for openEXR format files.

The lowest subband is recorded as HDR floating-points, since the proposed lifting introduced non-linearity and quantization error will be exaggerated by TMO. Since only the transform differences between the proposed method and its counterpart are required to evaluate the decorrelation performance The above quantization method is adopted. The advanced quantization methods can be applied to improve the performance, such as optimal bit allocation.

3.3.2.3 Entropy Measurement

Theoretical Extropy

The theoretical bit-rate can be estimated from the entropy of quantized subbands:

$$H = - \sum_{i=1}^n p(x_i) \log p(x_i), \quad (3.99)$$

where p is the probability of each unique pixel value x_i , n is the total number of unique pixel value,

$$H_a = \sum_{s=1}^n \frac{H_s}{R}, \quad (3.100)$$

where H_a is the average entropy of the whole image, H_s is the entropy of each DWT subband, n is the total number of DWT subbands, result in bit per pixel. The entropy of an image illustrates the cost prediction of compression the picture – the least bit can describe this image without loss information. Firstly, the entropy of each subband in terms of all pixels in this subband is computed and then the summation of each subband's entropy is calculated. Entropy summation divided by the total numbers of pixels of the tested image is the compressed bit per pixel. The entropy is calculated by quantized 3-level wavelet pyramid matrix. The increase of quantization factors results in increasing in bit per pixel. As mentioned above, the additional compensation layers need to be calculated when applying 5/3 lifting.

Actual Entropy coding

The entropy coders of JPEG, JPEG2000 and H.264 are different (QM,

MQ, CABAC), they are all arithmetic coding. The current arithmetic coder – LzTurbo is selected as the actual entropy coder. The actual entropy encoding performance is around 50% of the theoretical cost.

3.3.3 Experimental Results and Discussion

The BD-PSNR [11, 12] computes the relative difference between two corresponding RD curves. Since the peak of the HDR content is usually unknown, and research suggests [3] that the HDR-VDP [88] metric is the closest visual quality metric to the subjective ground truth HDR. The BD-MSE and BD-PSNR are used instead of BD-PSNR. The BD-MSE and BD-HDR-VDP are computed by 5 RD points out of 11 points – from the 5th to the 9th for each method, the higher BD-MSE and BD-HDR-VDP value means higher average MSE and HDR-VDP improvement with respect to the 9/7-t method. The following Table 3.1 and Table 3.2 are the HDR-VDP, the MSE results and the information of test-sets are in Section 2.7.

The reference for BD results of above tables is 9/7-t [89], the t represents the traditional approaches which are tone-mapping first approaches. The p represents the proposed DWT-TMO methods. Tables 3.1 and 3.2 demonstrated each wavelet has superior HDR-VDP performance in different test pictures, and 9/7-p has overall better HDR-VDP results. The larger positive BD-VDP and BD-MSE results suggest that this method is better than the 9/7-t method, and negative results suggest otherwise.

The results of EXR 7 and 8 as displayed in Figure 3.8, where the HDR-VDP of proposed 5/3 method were not decreased to 0 when $Q_f = 50000$. This is because of the pixel values of the original image are not evenly distributed, and the quantization error will be amplified by ITMO. By increasing the Q_f , the decreasing trend of HDR-VDP and MSE results can be illustrated. And because of the LL subband of 5/3 method is a cross HDR and LDR subband, the MSE results have abrupt distortions in the first few low bit-rates, and MSE decreased eventually when the bit-rate increased.

Table 3.1: BD-HDR-VDP versus bit-rate results of test-set with respect to the reference 9/7-t – The larger positive BD-VDP and BD-MSE results suggest that this method is better than the 9/7-t method, and negative results suggest otherwise.

| no. | Haar-p | Haar-t | 3/5-p | 3/5-t | 5/3-p | 5/3-t | 9/7-p |
|---------------------------|--------|--------|-------|-------|-------|-------------|-------------|
| 1 RGBE | 0.32 | 0.17 | 0.03 | -0.12 | 0.25 | 0.16 | 0.2 |
| 2 | 0.29 | 0.12 | 0.14 | -0.06 | 0.27 | 0.06 | 0.19 |
| 3 | 0.02 | 0.01 | 0.02 | -0.07 | 0.02 | 0.01 | 0.02 |
| 4 | 0.24 | 0.2 | 0.22 | 0.1 | 0.23 | 0.19 | 0.17 |
| 5 | -0.49 | -0.12 | -0.56 | -0.35 | 0.07 | 0.05 | -0.04 |
| 6 | 0.1 | 0.08 | -0.1 | -0.13 | 0.13 | 0.05 | 0.08 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0.14 | 0.1 | 0.11 | -0.09 | 0.13 | 0.07 | 0.1 |
| 9 | 0.02 | 0.01 | 0.01 | -0.06 | 0 | 0.02 | 0.02 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | 0.56 | 0.22 | 0.22 | 0.04 | 0.07 | 0.09 | 0.24 |
| 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0.13 | 0.1 | -0.06 | -0.16 | 0.1 | 0.05 | 0.06 |
| 14 | -0.17 | 0.03 | -0.63 | -0.17 | 0.23 | 0.15 | 0.19 |
| 15 | -0.26 | -0.11 | -0.3 | -0.24 | 0.04 | 0.04 | 0 |
| 16 | 0.02 | 0.05 | -0.25 | -0.25 | 0.03 | 0.04 | -0.01 |
| 17 | 0.41 | 0.07 | 0.32 | -0.21 | 0.35 | 0.18 | 0.39 |
| 18 | 0.03 | 0.04 | -0.11 | -0.11 | 0.08 | 0.05 | 0.05 |
| 19 | -0.13 | 0.01 | -0.15 | -0.11 | 0.05 | 0.03 | 0.03 |
| 20 | 0.03 | 0.01 | -0.04 | -0.17 | 0.04 | 0.03 | 0.03 |
| 21 | 0 | 0 | 0 | 0 | -0.06 | 0 | 0 |
| 22 | -0.77 | -0.12 | -0.81 | -0.4 | 0.05 | 0.04 | -0.03 |
| 23 | -0.76 | -0.06 | -0.93 | -0.39 | 0.06 | 0.04 | 0.03 |
| 24 | -0.7 | -0.28 | -0.74 | -0.56 | 0.05 | 0.04 | -0.04 |
| 25 | 0 | 0 | 0 | -0.01 | -0.87 | 0 | 0 |
| 1 EXR | 0.01 | 0.01 | 0 | -0.03 | 0.01 | 0 | 0 |
| 2 | 0.09 | 0.02 | -0.03 | -0.31 | 0.11 | 0.07 | 0.08 |
| 3 | 0.02 | 0.01 | 0.01 | -0.06 | 0.02 | 0.01 | 0.02 |
| 4 | 0 | 0 | 0.01 | 0 | 0 | 0 | 0 |
| 5 | 0.13 | 0.08 | 0.02 | -0.15 | 0.1 | 0.07 | 0.08 |
| 6 | -0.09 | -0.05 | -0.1 | -0.1 | -0.08 | 0.02 | -0.02 |
| 7 | -0.68 | -0.25 | -0.9 | -0.42 | -0.78 | 0.04 | -0.06 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | -0.78 | -0.09 | -0.89 | -0.44 | 0.02 | 0.02 | -0.02 |
| mean [-0.3 0.3] | 0.02 | 0.03 | -0.06 | -0.08 | 0.02 | 0.04 | 0.05 |
| std [-0.3 0.3] | 0.13 | 0.06 | 0.18 | 0.09 | 0.24 | 0.05 | 0.06 |

Table 3.2: BD-MSE versus bit-rate results of test-set with respect to the reference 9/7-t – The larger positive BD-VDP and BD-MSE results suggest that this method is better than the 9/7-t method, and negative results suggest otherwise.

| no. | Haar-p | Haar-t | 3/5-p | 3/5-t | 5/3-p | 5/3-t | 9/7-p |
|---------------------------|--------|--------|---------|-------------|---------|-------|-------------|
| 1 RGBE | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | -20.81 | 0 | -65.13 | -0.07 | 0.01 | 0.01 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0.08 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0.02 | 9.3 | -0.32 | 11.07 | 4.05 | 4.06 | 0.09 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | 0 | 0 | 0 | 0 | -0.49 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | -0.01 | 0 | 0 | 0 | 0 |
| 15 | -862.8 | 0.04 | -1.77e4 | -1.46 | 0.21 | 0.16 | 0.32 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 21 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 22 | -0.45 | 0 | -7.29 | -0.03 | 0 | 0 | 0 |
| 23 | -0.7 | -0.19 | -226.3 | -1.74 | 0.3 | 0.19 | 0.22 |
| 24 | -2.52 | -0.01 | -104.54 | -0.06 | 0.01 | 0.01 | 0 |
| 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 EXR | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | -102.8 | 0.03 | -3.24e4 | -10.41 | 1.24 | 1.05 | 1.07 |
| 7 | -0.08 | 0 | -2.05 | 0 | -7.53e4 | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | -9.8 | 0 | 0 |
| 9 | -0.03 | 0 | -1.28 | 0 | 0 | 0 | 0 |
| mean [-100 100] | -1.01 | 0.37 | -16.81 | 0.38 | 0.16 | 0.17 | 0.01 |
| std [-100 100] | 4.24 | 1.9 | 50.92 | 2.3 | 0.83 | 0.82 | 0.04 |

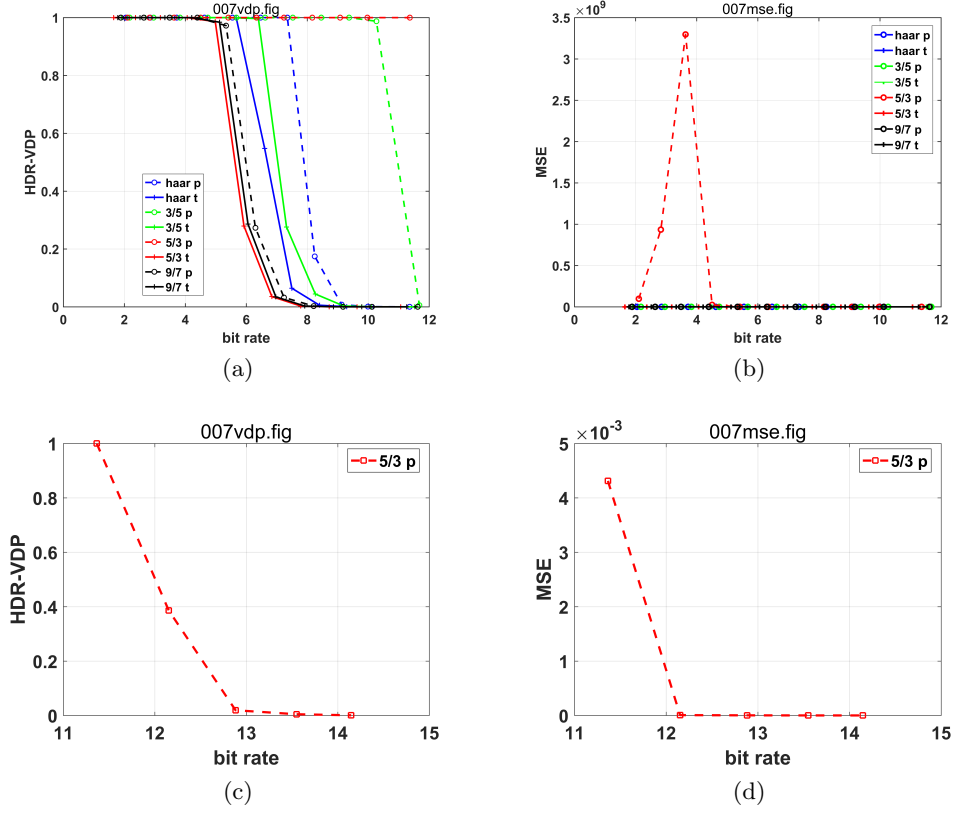


Figure 3.8: Results of Increasing Q_f of 5/3-p of EXR 007 (a) HDR-VDP, (b) MSE, (c) HDR-VDP of increased Q_f , (d) MSE of increased Q_f

The quality differences among Haar, 3/5, 5/3 and 9/7 liftings are noticeable under the same bit-rate, because of the characteristics of the different wavelets *i.e.* vanishing moment. It means a fine selection of wavelet can lead to a higher coding gain. The compression quality can be determined by the HDR-VDP metric with given view conditions. The proposed scheme can be extended to HDR video compression, and the flickering artifacts caused by TMO can be reduced by a temporal filtering of TMO brightness parameters.

Note: The complete RD results of all dataset are presented in the Appendix B, the mean and the standard deviation are limited to $([-0.3 \ 0.3])$ and $([-100 \ 100])$ respectively) because of the values of the extreme cases are too large.

The results suggest that the compression performances of different liftings are depended on the characteristics of the image. The longer filter length

and more vanishing moment of wavelets are not always outperforming their counterparts, especially in the extremely high contrast scenarios. Therefore The finer wavelet selection leads to a better visual quality of determined bit-rate. For some cases *e.g.* no. RGBE 5, 10, EXR 6 and 7, the BD-MSE are extremely high is due to the MSE is not decreased to 0 when the fixed quantization factor is applied $Q_f = 50000$. The MSE is decreased eventually when the bit-rate increased further the decreasing trend of HDR-VDP and MSE results is illustrated in the Figure3.8. Furthermore, the results suggested that the differences in HDR-VDP is equivalent to the differences in MSE, even though the MSE is not HVS driven metric.

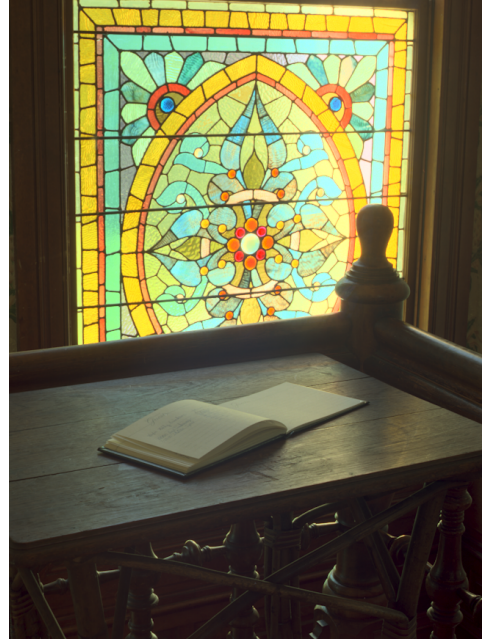
In some extreme cases *e.g.* no. EXR 6, 7, the 9/7 lifting performs worse than other liftings because of the pixel coherence of HDR image is less than the LDR image, the longer filter length could not predict and update the HDR signal precisely. The shortcoming of the proposed liftings is the prediction and decoding steps require a LDR/HDR mixed subband, the quantization errors are relatively greater than the other methods for some extreme cases *e.g.* no. EXR 6, 7. The negative coefficients of low-pass filters *e.g.* 5/3 and 9/7 introduce the negative coefficients in the LL subband, which affect the TMO for LL subband.

Side by Side Comparisons of Different liftings

The following Figures 3.9 to 3.14 show the side by side results of the proposed methods and traditional methods. The orders of examples are following lower HDR-VDP detection rate method, higher HDR-VDP detection rate method, the original image, HDR-VDP detection probability map of the first image and the second image. The visual differences predicted by HDR-VDP metric are due to compression distortions. The corresponding RD HDR-VDP results are displayed in Figure 3.15, the results confirmed that the distortions among Haar, 3/5, 5/3 and 9/7 liftings are noticeable under the same bit-rate, it means a better wavelet decomposition of the image can lead to a higher coding gain.

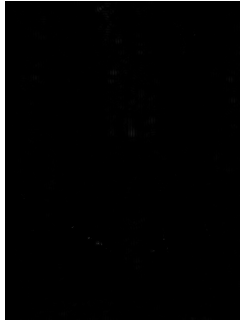


(a)

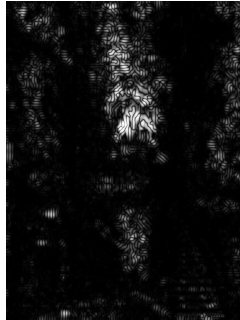


(b)

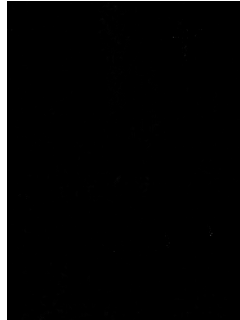
Figure 3.9: (a) Original image of RGBE002, (b) original image of RGBE005.



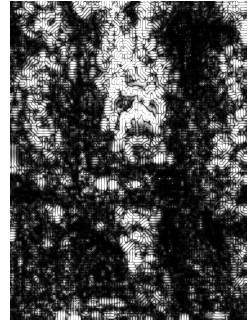
(a)



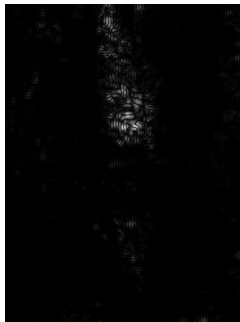
(b)



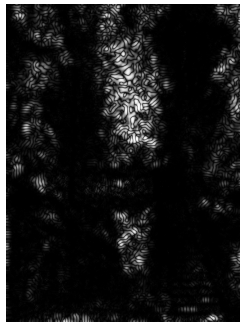
(c)



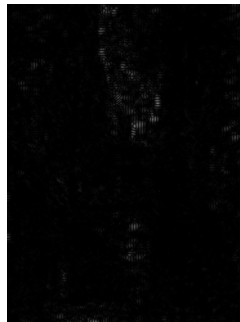
(d)



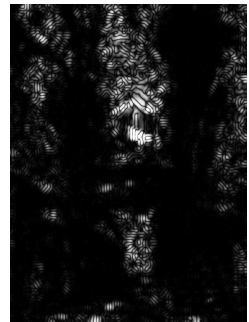
(e)



(f)



(g)



(h)

Figure 3.10: HDR-VDP maps of RGBE002 ($Q_f = 256$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t

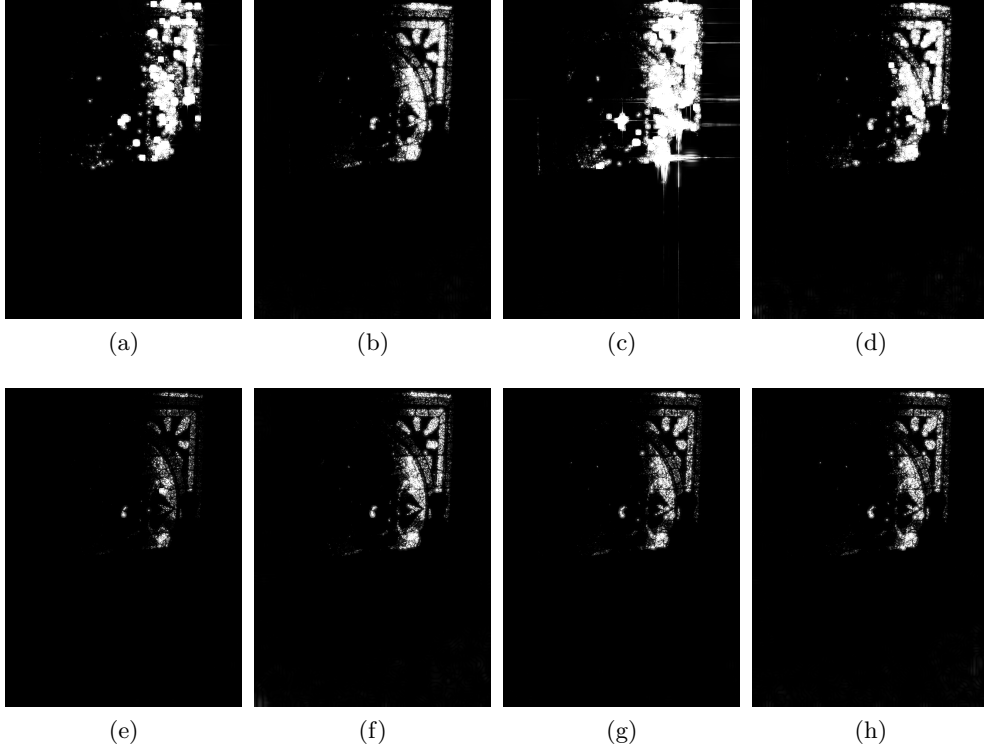


Figure 3.11: HDR-VDP maps of RGBE005 ($Q_f = 256$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t, (e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t

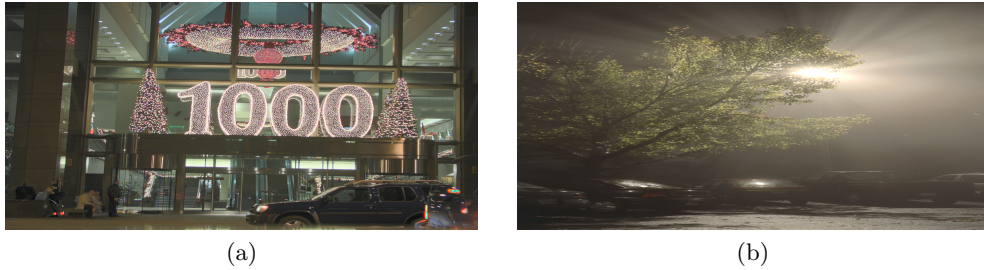


Figure 3.12: (a) original image of RGBE006, (b) original image of RGBE017.

3.3.4 Extending Propose DWT-TMO Lifting to HDR Video

Since the proposed DWT-TMO lifting method is for spatial compression only, it can be extended to motion video codings such as Motion JPEG and Motion JPEG2000 standards frame by frame. The TMOs introduced temporal artifacts, which can be reduced by filtering TMO parameters to ensure no sudden changes in brightness.

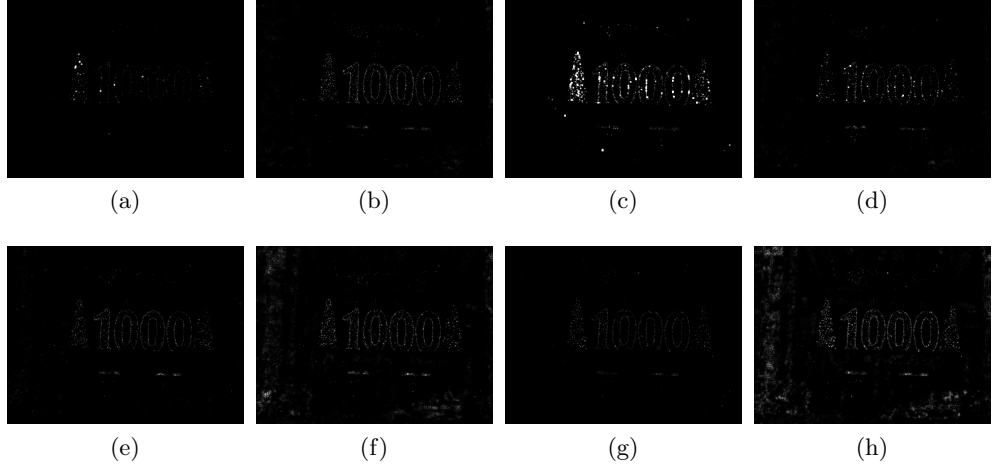


Figure 3.13: HDR-VDP maps of RGBE017 ($Q_f = 150$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t

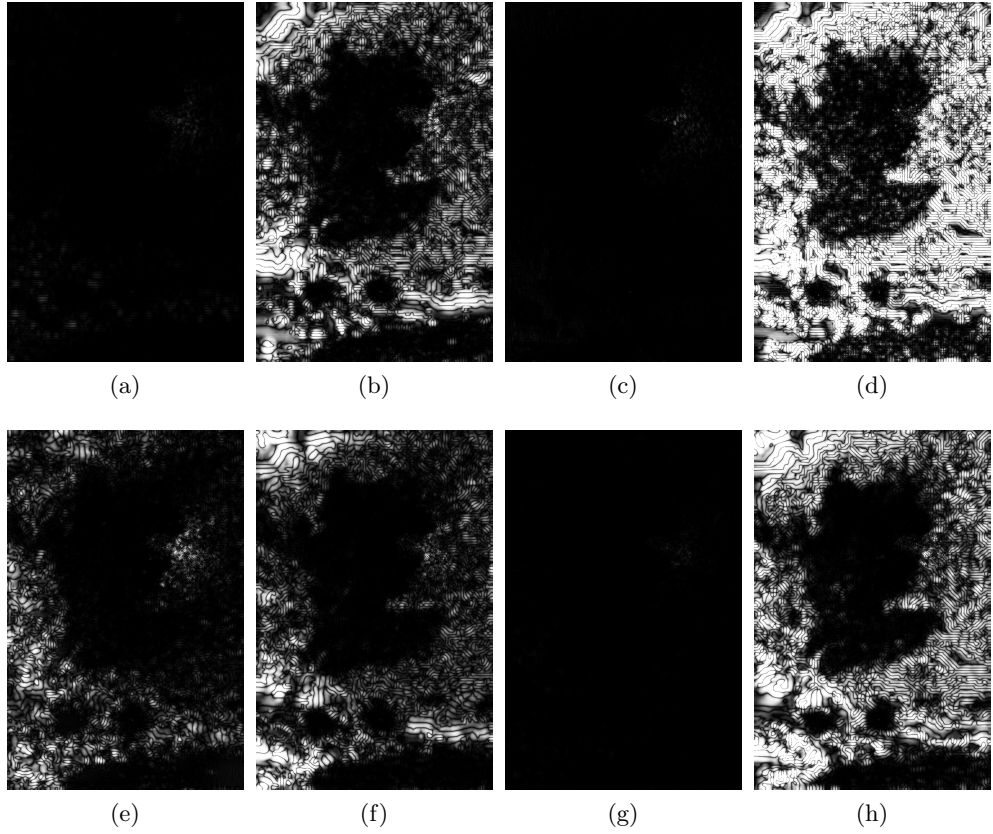


Figure 3.14: HDR-VDP maps of RGBE009 ($Q_f = 150$) (a)Haar-p, (b)Haar-t, (c)3/5-p, (d)3/5-t,(e)5/3-p, (f)5/3-t, (g)9/7-p, (h)9/7-t

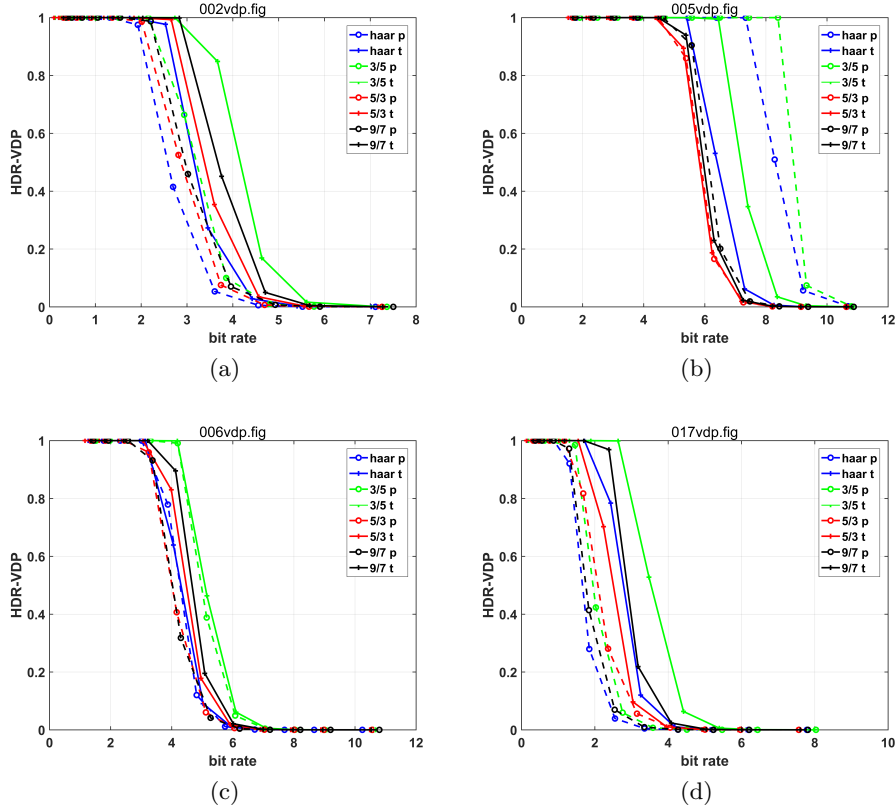


Figure 3.15: Results of HDR-VDP versus bit-rate (a) RGBE002, (b) RGBE005, (c) RGBE006, (d) RGBE017.

3.3.4.1 Tone mapped HDR Video compression artifacts

The temporal filtering [65] of the parameters of the TMO reduces the rapid frame to frame changes. The update equations to compute the filtered A_n , B_n , and a_n for the n^{th} frames are given by:

$$A_n = (1 - \alpha_A)A_{n-1} + (\alpha_A)A, \quad (3.101)$$

$$B_n = (1 - \alpha_B)B_{n-1} + (\alpha_B)B, \quad (3.102)$$

$$a_n = (1 - \alpha_a)a_{n-1} + (\alpha_a)a, \quad (3.103)$$

The parameters α_A , α_B , and α_a are time constants $\in [0, 1]$ and all are set to 0.98, it is because of this ensures the most – 0.98 of current frame weight is from the last frame, and only 0.02 of current frame’s weight is coming from

the current frame.

The parameter estimation requires the minimum and maximum luminance of the image L_{min} and L_{max} . With L_{av} the geometric mean of the image, the key of the scene a is estimated as:

$$a = 0.18 * 2^{2(B-A)/(A-B)}, \quad (3.104)$$

$$A = L_{max} - L_{av}, \quad (3.105)$$

$$B = L_{av} - L_{min}. \quad (3.106)$$

Photographic tone reproduction maps the geometric mean of the image to the estimated key value,

$$L(x, y) = aL(x, y)/L_{av} \quad (3.107)$$

before applying a non-linear compression:

$$L_d(x, y) = \frac{L(x, y)(1 + \frac{L(x, y)}{L_{white}^2})}{1 + L(x, y)}, \quad (3.108)$$

where L_{white} is a user parameter to specify the smallest input value that should be mapped to white.

As described in [36], the global flickering artifacts occurred in the Desert sequence from (105-106 frames):

<http://people.irisa.fr/Ronan.Boitard/>

However, temporal filtering methods have to trade-off between temporal and spatial contrast, since the high local contrast will be filtered to avoid abrupt temporal contrast changes.

The TMOs change the dynamic range of the HDR contents, and the quantization errors are affected accordingly. The quantization errors are then further increased and decreased correspondingly when decoding the HDR content. Beside of distortions, the TMOs also introduced temporal artifacts. To

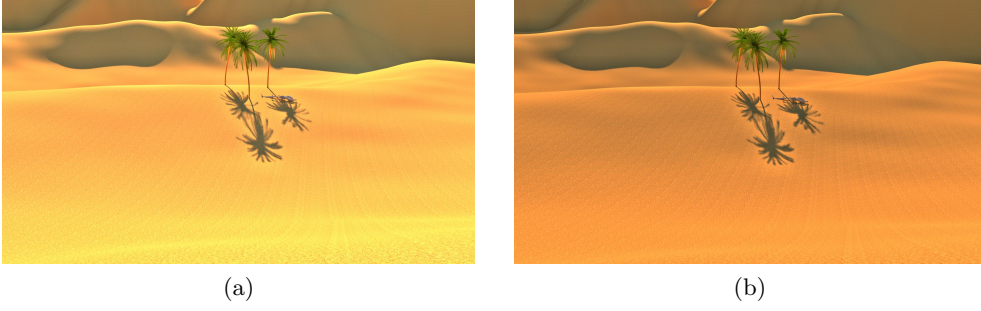


Figure 3.16: (a) Tone mapped No. 105 frame of Desert sequence; (b) Tone mapped No. 106 frame of Desert sequence.

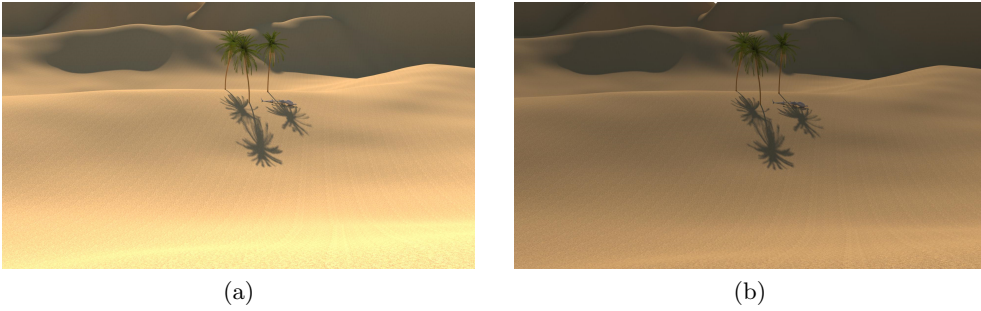


Figure 3.17: (a) Temporal filtered Tone mapped No. 105 frame of Desert sequence; (b) Temporal filtered Tone mapped No. 106 frame of Desert sequence.

avoid these limitations, the structure of HDR formats is reviewed, and the importance of the TMO in the HDR compression is evaluated in the next Chapter.

3.4 Summary

This chapter explored the relationship of DWT lifting steps and TMO. A novel compression scheme for HDR images based on DWT lifting with TMO built-in is proposed. The demonstration starts from 1D Haar lifting, and extends to 2D Haar, 2D $3/5$, 2D $5/3$ and 2D $9/7$ liftings. The properties of global TMO and ITMO integrating to the lifting are also introduced. This approach achieved a comparable performance in MSE and HDR-VDP of existing HDR image compression methods, with LDR backwards-compatibility. Under the same bit-rate, the quality differences among Haar, $3/5$, $5/3$ and $9/7$ liftings are

noticeable, because of the characteristics of the different wavelets *i.e.* vanishing moment. It means a fine selection of wavelet can lead to a higher coding gain. The compression quality can be determined by the HDR-VDP metric with given view conditions. The proposed scheme can be extended to HDR video compression, and the flickering artifacts caused by TMO can be reduced by a temporal filtering of TMO brightness parameters.

Based on the evaluation results of this chapter and with the development of HDR displays, a low complexity HDR compression scheme without LDR backwards-compatibility is proposed in the next Chapter. Because the artifacts caused by TMOs only affect the LDR sequences. The LDR backwards-compatibility can be obtained by applying TMO to the fully decoded bit-stream. The development of HDR displays will reduce the requirements of LDR backwards-compatibility. Furthermore, the high-levels of HDR compressions are more efficient without changing the legacy codec *i.e.* JPEG-XT, HEVC, although JPEG2000 based methods are better in compression performance. These will be addressed in the next chapter.

Chapter 4

Mapping-Free HDR Visual Content Compression

The importance of the TMO in the HDR compression is evaluated in this chapter. A mapping-free post HDR image compression based on JPEG and JPEG2000 standard codecs for current HDR image formats is proposed. This approach exploits the structure of HDR formats. It has equivalent compression performance and lowest computational complexity compared to the existing HDR lossy compressions.

4.1 Motivation

The RGBE, EXR and PFM HDR image formats [107] are mainly concerned on lossless compression for the perfect scene-referred image reconstruction. However, these formats are not efficient because they lack the consideration of the HVS limitations. Therefore, lossy HDR image compression alternatives like JPEG-HDR [131] and JPEG2000-HDR [135] have been proposed. In these methods, the dynamic range is compressed first to meet the input requirements of conventional compression codecs. For example, the dynamic range of the input HDR raw image is reduced by introducing a TMO [108, 37, 84] and the HDR information is recorded by storing the HDR residual layer in

JPEG-HDR. In contrast, since the JPEG2000 codec supports more bit-depth than the JPEG codec, the HDR dynamic range is reduced by an approximate logarithmic function, followed by 16-bit per colour channel in JPEG2000 compression by JPEG2000-HDR. These methods require binary to floating-point conversion and a mapping step before feeding into the JPEG and JPEG2000 codecs. Moreover, the TMOs change the dynamic range of the HDR contents, and the quantization errors are affected accordingly. The quantization errors are then further increased and decreased correspondingly when decoding the HDR content.

This chapter presents a mapping-free post HDR image compression based on JPEG and JPEG2000 standards for lossless HDR image formats. The term 'mapping' refers to the integer to floating-point conversion which happened during decoding RGBE and EXR image to HDR image, and HDR to LDR operation usually with a TMO. Recent TMOs claim that they preserve most of the perception of the HDR contents. mapping-free means that the reconstruction is always on the HDR scene-referred domain, and there is no distortion or loss of information due to TMO. The results in [3] illustrate that the HDR-VDP [88] metric is the closest visual quality metric to the subjective ground truth HDR. The JPEG-XT standard [3] is the official extension of JPEG to support HDR image encoding. On this basis, the JPEG2000-XT with the same structure as JPEG-XT is proposed and the performance is evaluated with mapping-free compression method together.

4.2 Characteristics Of Existing HDR Image Formats

The three most widely used lossless HDR image formats are: RGBE, EXR and PFM. The research [134] suggested that the relative step size must keep under 1% to prevent visible luminance quantization effect. This is the reason why all relative step sizes of HDR encodings satisfy such a requirement.

The RGBE is the most commonly adopted HDR format in published papers and available online HDR sources, OpenEXR follows next. The state of the art HDR compression algorithms usually referenced RGBE and OpenEXR images as test-set. Most HDR compressions operate in the scene-referred HDR domain [131, 135, 94, 142], which are the result of converting RGBE or OpenEXR to floating-point value, rather than computing the binary values of these two formats directly. The motivation is to exploit the existing LDR image encoding codecs directly without any modification, see Figures 4.1 and 4.2. The HDR binary values can be fed into the legacy codecs directly without integer to floating-points mapping, TMO, ITMO or logarithmic function. The LDR backward-compatibility can be obtained by applying TMO to the fully decoded bitstream. The blocks of dashed line in Figures 4.1 and 4.2 can be avoid in the proposed method.

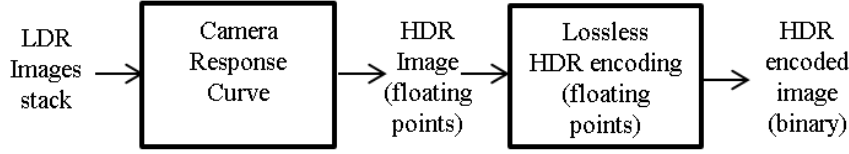


Figure 4.1: The usual HDR acquisition block diagram.

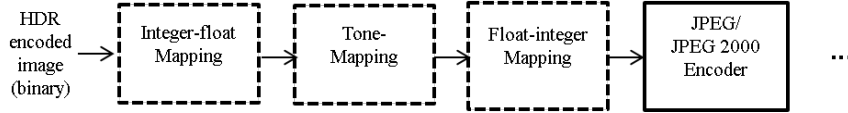


Figure 4.2: The existing HDR encoding block diagram.

The following sections introduced the proposed mapping-free methods and the proof of the concept in detail.

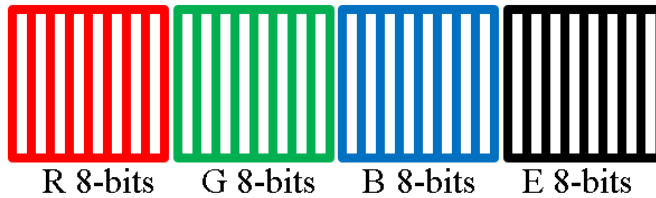


Figure 4.3: The binary representation of RGBE format.



Figure 4.4: The binary representation of EXR format.

4.3 Proposed mapping-free Post HDR Image Compression Method

The proposed mapping-free post HDR image compression and decompression schemes for JPEG/ JPEG2000 are shown in Figure 4.5. The input of mapping-free post HDR image compression is an integer value for the RGBE and OpenEXR (half-precision) image formats. Since the core codecs of JPEG and JPEG2000 remain unchanged, it requires less computational complexity than JPEG-HDR and JPEG2000-HDR. To be specific, all TMO ITMO related blocks in Figure 2.9, Figure 2.11 and Figure 2.13 are skipped in the proposed method. The mapping-free post HDR image compression based on JPEG for OpenEXR (half-precision) is not possible since the maximum encoding depth of JPEG2000 is restricted to 16-bit.



Figure 4.5: The proposed mapping-free HDR content compression block diagram.

4.3.1 Computational Complexities

The proposed mapping-free HDR compression scheme reduces the total complexity of existing methods. By removing the TMO and ITMO ($nO(M(1)k)$), division and multiplication ($nO(1)$) the computational complexities of the compression is recuded. Where n is the total number of pixels, M is the complexity of the chosen multiplication or division algorithm, k is the exponent of

TMO and ITMO. The detailed computational complexities are presented in the evaluation section.

4.3.2 Proof of the concept

Most of the HDR contents are stored as single or half-floating-point representations.

The relative errors between double-precision floating-point and single-precision floating-point formats conversion cannot be detected by HVS [107]. The following Figure 4.6 demonstrated the errors between single-precision floating-point and half-precision floating-point formats conversion. This proves that

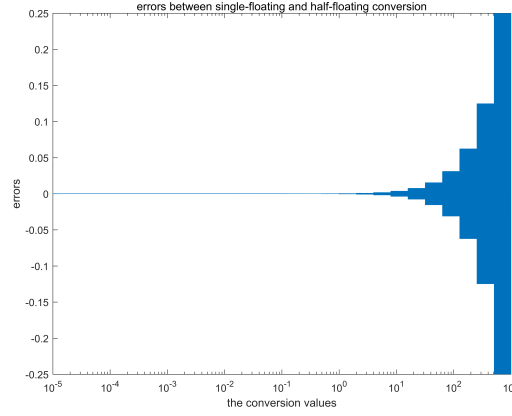


Figure 4.6: Errors between single-precision floating-point and half-precision floating-point formats conversion.

the errors occurred by conversion single-precision floating-point and half-precision floating-point formats are always below the JND and the HVS discrimination threshold, since the relative steps size is smaller than 1%. Therefore, the half-floating-point format can be used to record the HDR values, and they are stored as the corresponding 16-bit integers.

The converting of HDR floating-points to 16-bit integer is based on the logarithmic (inverse exponential) function:

$$I = (-1)^s * 2^{(e-15)} * 1.m, \quad (4.1)$$

which is the inverse function of this equation. Where s is sign bit, e is exponential bits and m is mantissa bits.

The TMOs also tend to be the logarithmic functions which approximate the HVS. Figure 4.7 shows the tone-mapping function, the IEEE floating-points transform function of given HDR image and its tone mapped version. For a given HDR image, the corresponding histograms are displayed in Figure 4.8

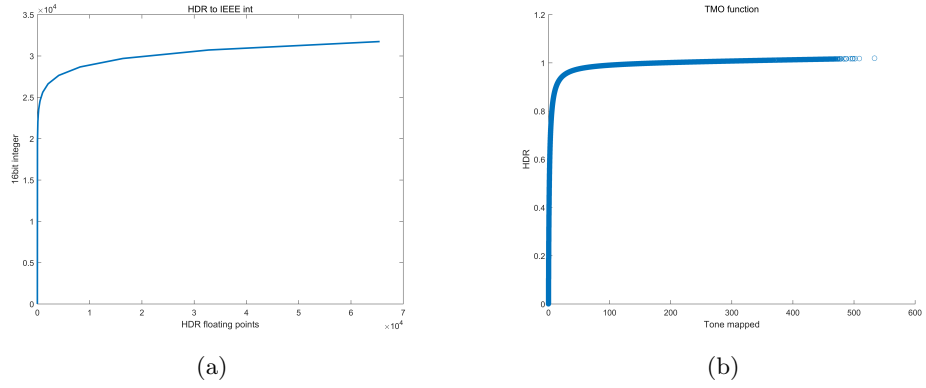


Figure 4.7: (a)IEEE floating to integer conversion, (b)TMO function

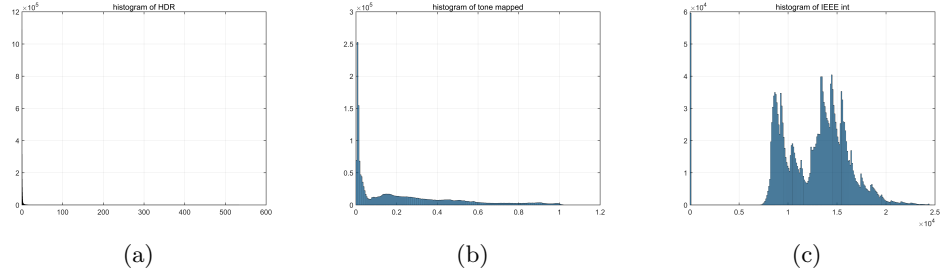


Figure 4.8: (a)Histogram of HDR, (b)Histogram of Tone mapped HDR, (c)Histogram of IEEE half-floating integer.

To conclude, when the HDR points were stored as IEEE floating-points representations, a logarithmic function is automatically applied on HDR pixels. Hence, the none tone-mapped integers could be the input of compression codecs and obtain similar compression performance compared to tone mapping method. The range of HDR image is usually greater than HDR videos because the images are a stack of several exposures at least from 8-bit 5 stacks

Table 4.1: Evaluation of Transform

| Seq no. | Entropy | | | | | Correlation | | | |
|------------|---------|-------|--------|--------|--------|-------------|--------|--------|--------|
| | HDR | MF | 10-bit | 12-bit | 16-bit | MF | 10-bit | 12-bit | 16-bit |
| 1 | 14.63 | 12.75 | 7.53 | 9.43 | 12.86 | 0.297 | 0.542 | 0.542 | 0.542 |
| 2 | 14.63 | 12.75 | 7.53 | 9.43 | 12.86 | 0.297 | 0.542 | 0.542 | 0.542 |
| 3 | 14.63 | 12.58 | 7.03 | 8.93 | 12.4 | 0.389 | 0.627 | 0.627 | 0.627 |
| 4 | 14.56 | 12.2 | 7.57 | 9.48 | 12.97 | 0.385 | 0.642 | 0.642 | 0.642 |
| 5 | 14.63 | 12.86 | 6.82 | 8.73 | 12.23 | 0.238 | 0.582 | 0.582 | 0.582 |
| 6 | 14.61 | 12.69 | 4.95 | 6.83 | 10.42 | 0.053 | 0.755 | 0.755 | 0.755 |
| 7 | 14.63 | 12.75 | 7.13 | 9.04 | 12.56 | 0.34 | 0.544 | 0.544 | 0.544 |
| 8 | 14.59 | 12.35 | 7.28 | 9.18 | 12.64 | 0.287 | 0.695 | 0.695 | 0.695 |
| 9 | 14.58 | 12.46 | 7.06 | 8.95 | 12.42 | 0.265 | 0.597 | 0.597 | 0.597 |

and the videos are generated by 12-bit shifted by 8-bit 2 stacks.

The mean of entropies and correlation of original HDR, IEEE half-integer mapping, 10-bit, 12-bit and 16-bit perceptual quantization of BT 2020 standard mapping are displayed in Table.4.1, the sequences are listed in Section 2.7.3: The results illustrate that the IEEE half-floating integer representation offers better decorrelation than perceptual quantization method.

A small change in exponent value results in a huge magnitude change in scene referred pixel. The above experimental results demonstrated that the exponent channel has to be compressed in lossless mode in both JPEG and JPEG2000 to recover the proper HDR scene-referred image.

The OpenEXR supports up to 32-bit per single channel, 16-bit half data type is the dominating standard. The error tolerance of the half-precision representation is closing to the discrimination threshold of HVS and reaching the precision limitation of sensors and ADC [107]. The half data type is a logical contraction of the IEEE-754 floating-point representation to 16-bit [1].

The direct compression based on JPEG for OpenEXR half-precision is not possible since the single input layer is 16-bit depth. In this chapter, the direct JPEG2000 compression for OpenEXR are evaluated. The 16-bit binary value can be input to JPEG2000 codec directly without converting to floating-point and mapping to integer steps.

4.3.3 Comparison to Perceptual based Quantization

The 10 and 12-bit RGB to YCbCr transforms of ITU-R BT 2020 standard (JCT-VC-w15793 meeting) [104, 116, 126] are:

$$(R, G, B)' = T(\max(0, \min((R, G, B)/10000, 1))), \quad (4.2)$$

where

$$T(R) = \left(\frac{c_1 + c_2 R^{m_1}}{1 + c_3 R^{m_1}} \right)^{m_2}, \quad (4.3)$$

where

$$\begin{aligned} m_1 &= 0.1593017578125, \\ m_2 &= 78.84375, \\ c_1 &= c_3 - c_2 + 1 = 0.8359375, \\ c_2 &= 18.8515625, \\ c_3 &= 18.6875. \end{aligned} \quad (4.4)$$

They are constants in the standard, therefore the precision is unchanged. The YCC are converted by BT.2020 standard:

$$\begin{aligned} Y' &= 0.262700 * R' + 0.678000 * G' + 0.059300 * B', \\ Cb' &= -0.139630 * R' - 0.360370 * G' + 0.500000 * B', \\ Cr' &= 0.500000 * R' - 0.459786 * G' - 0.040214 * B'. \end{aligned} \quad (4.5)$$

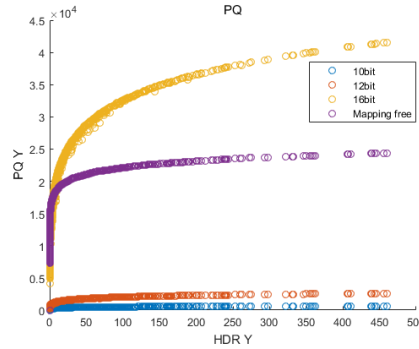


Figure 4.9: (a) 10, 12, 16-bit PQ transform and IEEE floating to integer conversion

Table 4.2: The upper quality limits

| Bit depth | HDR-VDP | HDR-VQM |
|--------------|---------|---------|
| 8 | 6.82 | 8.79407 |
| 10 | 6.85 | 8.75501 |
| 12 | 6.98 | 8.75146 |
| 14 | 7.59 | 8.75136 |
| 16 | 7.66 | 8.75133 |
| Best Quality | 100 | 0 |

The 10, 12 and 16-bit RGB to YCbCr transform in Figure 4.9 are followed ITU-R BT 2020 standard. Figure 4.10 shows the rate-distortion results of the described method.

The results show that the PQ method is designed for 10-bit for optimal performance. The cost of 16-bit increases while the quality remains the same, the green and the yellow RD curves are almost identical. To further explain this, the upper quality limits of 8, 10, 12, 14 and 16-bit PQ method are calculated respectively, and they are displayed in Table4.2. The HDR-VDP score used is Mean-Opinion-Score (MOS) prediction because the final prediction probability are the same which are 1 for all the cases. The factors of PQ method is designed low bit depth cases 8 or 10, the quality is not scaled proportional with the increased bit-depth, 12, 14 and 16-bit.

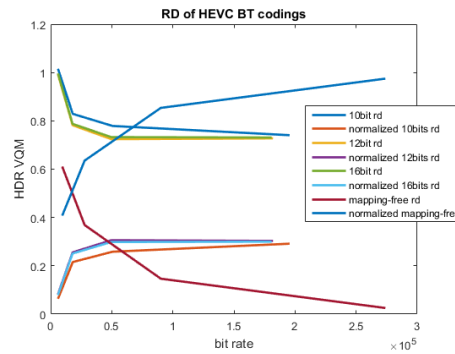


Figure 4.10: Result of HEVC codings.

4.4 Proposed JPEG2000-XT HDR Image Compression Method

The JPEG-XT is the new standard for compressing HDR image [3], and the data set is provided by [66]. The JPEG2000-XT compression method was inspired by the JPEG-XT standard profile-C.

Since JPEG2000 codec performs better than JPEG, the proposed JPEG2000-XT replaced JPEG codec in the JPEG-XT to JPEG2000. The bit depth of the two layers coding can be either 8 or 16. Figure 4.11 demonstrated that the closed-loop performs better than the open loop for 8-bit encoding, and open-loop outperforms closed loop for 16-bit coding. Because the 8-bit depth is not enough to record original 16-bit HDR information and 16-bit closed loop is over compensation.

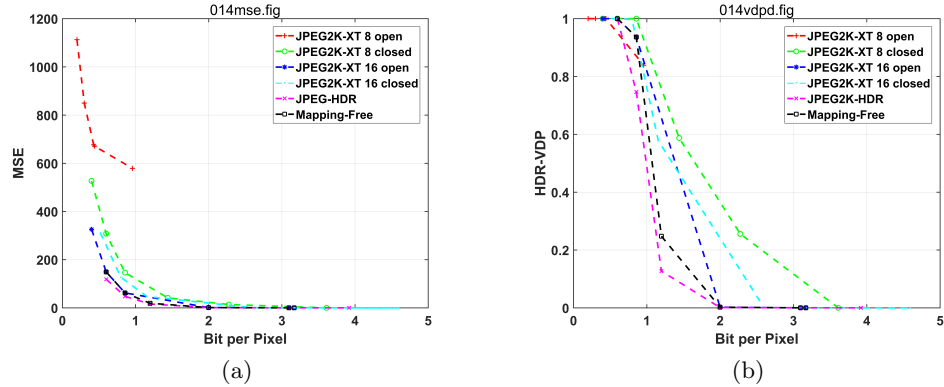


Figure 4.11: Open-loop and closed-loop performance of JPEG2000-XT compression (a) The MSE results of RGBE 014 image; (b) The HDR-VDP results of RGBE 014 image .

4.5 Proposed mapping-free Post HDR Video Compression Method

To extend the proposed work of [38], the Motion JPEG and Motion JPEG2000 codecs are tested for HDR videos which are provided by [42]. The compression categories of HDR video sequence are the same as the proposed HDR image

compression.

4.5.1 JPEG/ JPEG2000

The proposed JPEG/ JPEG2000 video compression scheme is to replace the block of JPEG and JPEG2000 codecs to Motion JPEG and Motion JPEG2000 codecs in Figure 4.5.

4.5.2 JPEG2000-XT

The proposed JPEG2000-XT video compression scheme is to replace the block of JPEG and JPEG2000 codecs in Figure 2.13

4.5.3 HEVC

The current HEVC version support up to 16-bit extended input depth, which is called as 'RExt' profile. The HDR video frame can be compressed directly by the proposed mapping-free HDR visual content method.

4.6 Result and Discussion

In this section, the results of the proposed method are evaluated and compared with the current state of the art methods.

4.6.1 Experimental Setups

The experiments are conducted on the image and video test-sets are introduced in Section 2.7. Each video sequence of a scene is selected.

4.6.2 Objective Quality Evaluation Metric

The same quality metrics are used as in the previous Chapter.

4.6.3 Experimental Results and Discussion

This section demonstrated the results of the proposed image and video compressions.

4.6.3.1 Computational Complexities' evaluation

The summation of encoding and decoding computational complexities of the test methods are shown in Table 4.3. The complexities of encoding and decoding process of JPEG and JPEG2000 are assumed as equal to $J(n)$. The complexities of a global TMO and ITMO are $nO(M(1)k)$, division and multiplication are $nO(1)$. The n is the total number of pixels, M is the complexity of the chosen multiplication/division algorithm, k is the exponential function of TMO and ITMO.

The proposed method has the lowest computational complexities compared with other methods when the desired output is HDR.

Table 4.3: Computational Complexities (per frame)

| Compressions | LDR | HDR |
|------------------|------------------------------|------------------------------|
| JPEG-HDR | $2nO(1) + nO(M(1)k) + 3J(n)$ | $3nO(1) + nO(M(1)k) + 4J(n)$ |
| JPEG-Xt | $2nO(M(1)k) + 4J(n)$ | $3nO(M(1)k) + 5J(n)$ |
| MF-JPEG | $nO(M(1)k) + 2J(n)$ | $2J(n)$ |
| JPEG2K-HDR | $3nO(M(1)k) + 2J(n)$ | $2nO(M(1)k) + 2J(n)$ |
| MF-JPEG2K | $nO(M(1)k) + 2J(n)$ | $2J(n)$ |

The actual computational time is computed by a i5-4690K 3.5GHz CPU, 8GB RAM, MATLAB 2016b. Where on the average TMO takes 0.32 secs. ITMO, division and multiplication take 0.02 secs. The JPEG and JPEG2K encoding take 0.02 secs and 0.1 secs respectively. The real computational time of these method are displayed in Table4.4.

4.6.3.2 Image results

In Figure 4.12, the rate-distortion (RD) curves of three images of the RGBE test-set – (Apartment.hdr, MtTamWest.hdr and SpheronSiggraph2001.hdr) are presented.

Table 4.4: Computational time (secs per frame)

| Compressions | LDR | HDR |
|------------------|------|------|
| JPEG-HDR | 0.96 | 1.18 |
| JPEG-Xt | 1.44 | 1.96 |
| MF-JPEG | 0.72 | 0.4 |
| JPEG2K-HDR | 2.96 | 2.64 |
| MF-JPEG2K | 2.32 | 2 |

For the result of 'Apartment.hdr' image in Figure 4.12 (a), (b), the performance of the mapping-free JPEG2K method is worse than other counterparts in MSE, it is because the exponent channel of RGBE file is lossless compressed to avoid the artifact. In terms of HDR-VDP, the proposed mapping-free and JPEG2000-Xt results are better than their counterparts.

For the 'MtTamWest.hdr' image in Figure 4.12 (c), (d), the mapping-free post HDR image compression method has higher bit per pixel requirements to obtain the same quality in terms of MSE and HDR-VDP. The JPEG2000-Xt compression performed close to JPEG-HDR and JPEG2000-HDR.

For the results of 'SpheronSiggraph2001.hdr' image in Figure 4.12 (e) (f), the overall compression performance of the JPEG2000-HDR method is the best among all methods in terms of HDR-VDP. The performances of the JPEG2000-Xt and mapping-free methods are similar to JPEG-Xt. The mapping-free post HDR image compression method can achieve lower HDR-VDP detection rate and MSE at higher bit per pixel while the JPEG-HDR failed.

In Figure 4.13, the RD curves of three images of the EXR test-set – (Flower.exr, Ocean.exr and StarField.exr) are demonstrated.

For the 'Flower.exr' image in Figure 4.13 (a), (b), the proposed mapping-free JPEG2000 and JPEG2000-Xt compression methods performed close to JPEG2000-HDR in both of MSE and HDR-VDP. For JPEG based methods, the performance of JPEG-HDR and JPEG-Xt are close and worse than JPEG2000 based methods.

For the 'Ocean.exr' image in Figure 4.13 (c), (d), the mapping-free post

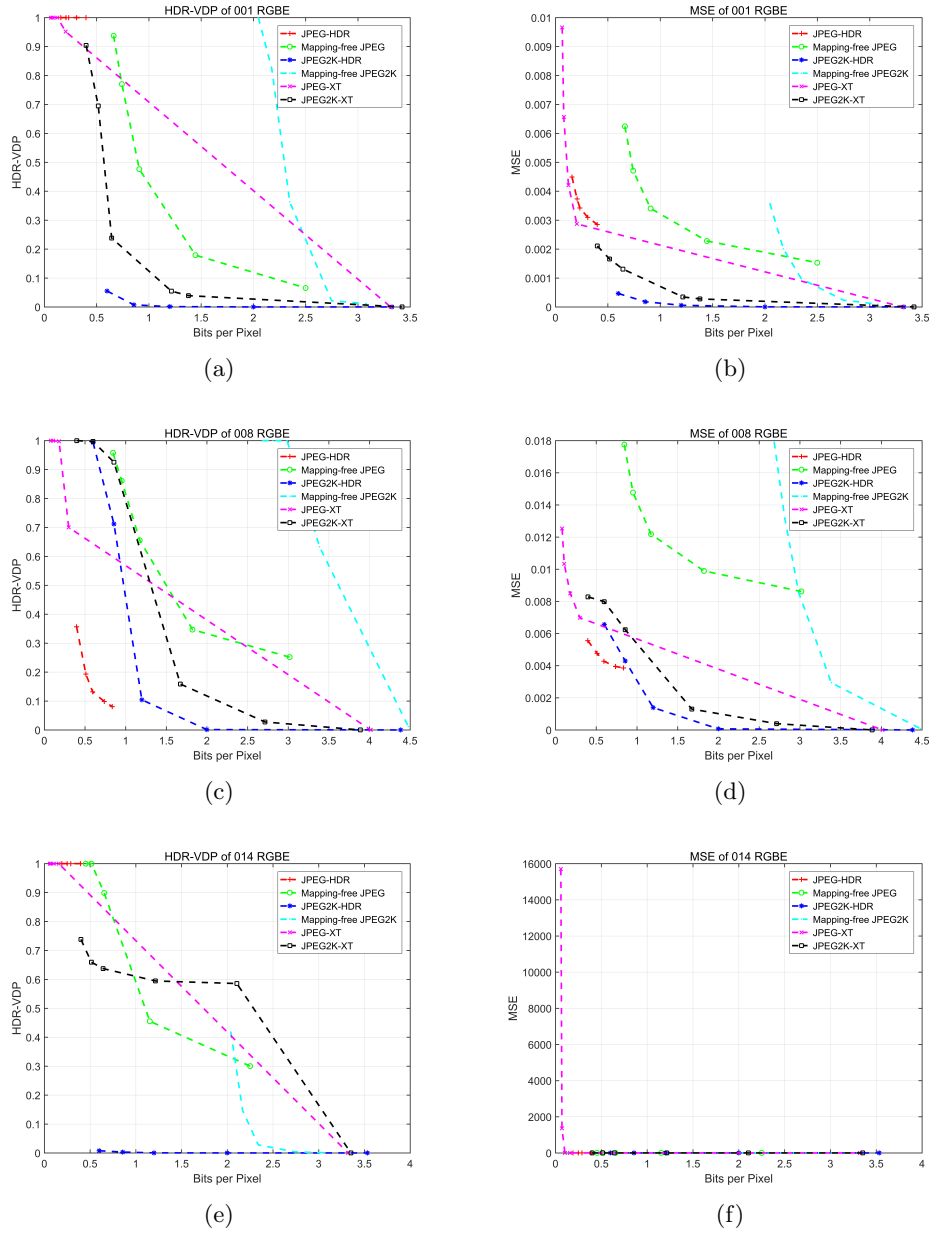


Figure 4.12: Rate-distortion curve of RGBE 001/ 008/ 014 images. (a) HDR-VDP result; (b) MSE result; (c) HDR-VDP result; (d) MSE result; (e) HDR-VDP result; (f) MSE result.

HDR image compression method has very close performance in terms of rate-distortion to JPEG2000-HDR and they have better compression under the same quality.

For the 'StarField.exr' image in Figure 4.13 (e), (f), mapping-free JPEG2000 method performed close to JPEG2000-HDR in both of MSE and HDR-VDP.

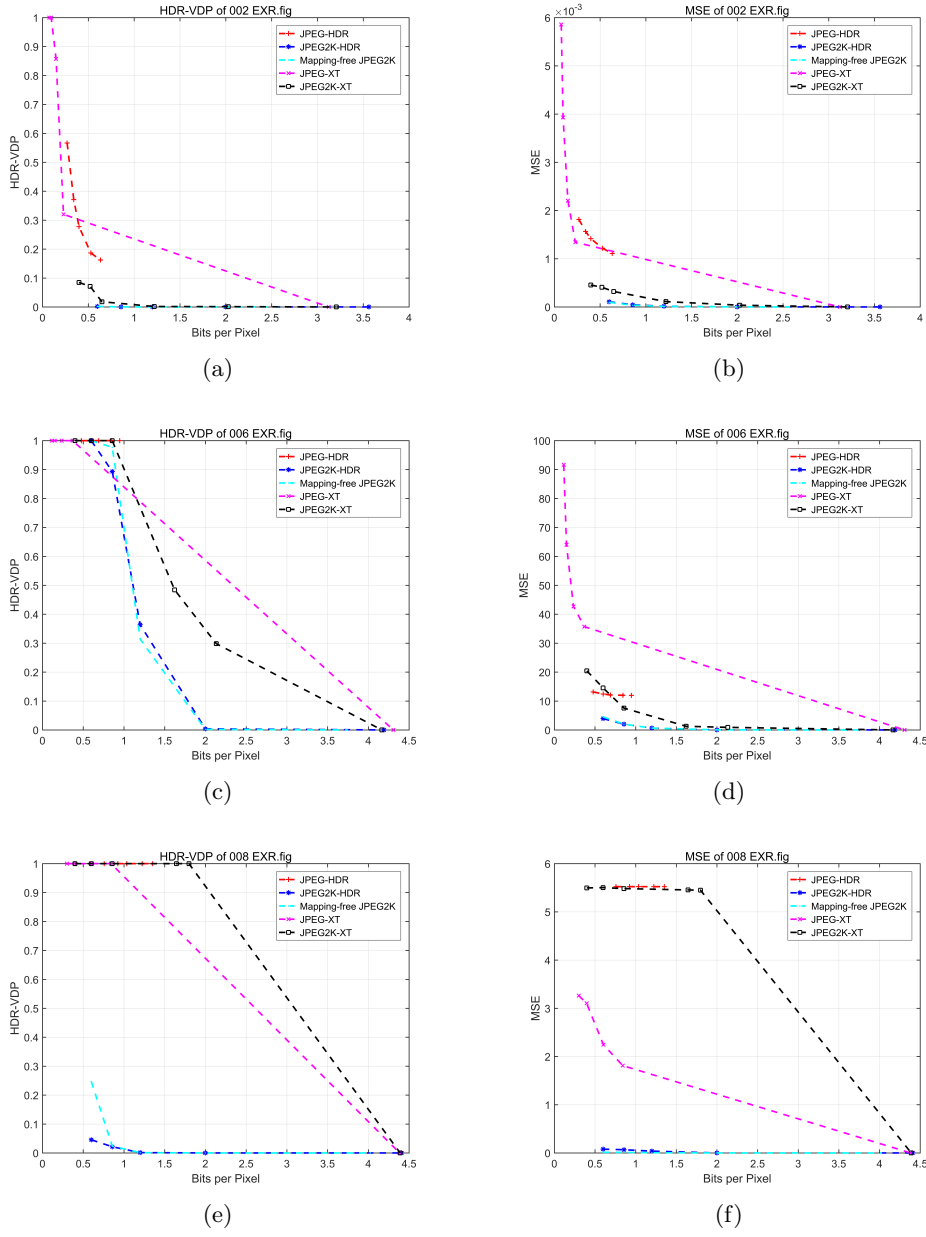


Figure 4.13: Rate-distortion curve of EXR 002/ 006/ 008 images. (a) HDR-VDP result; (b) MSE result; (c) HDR-VDP result; (d) MSE result; (e) HDR-VDP result; (f) MSE result.

The JPEG-Xt performs better than JPEG2K-Xt.

In Figure 4.14, the rate-distortion (RD) curves of three images of the PFM test-set – (McKeesPub.pfm, set22.pfm and set70.pfm) are presented.

For all of the results in Figure 4.14, the performance of mapping-free JPEG2000 method is close to JPEG2000-HDR.

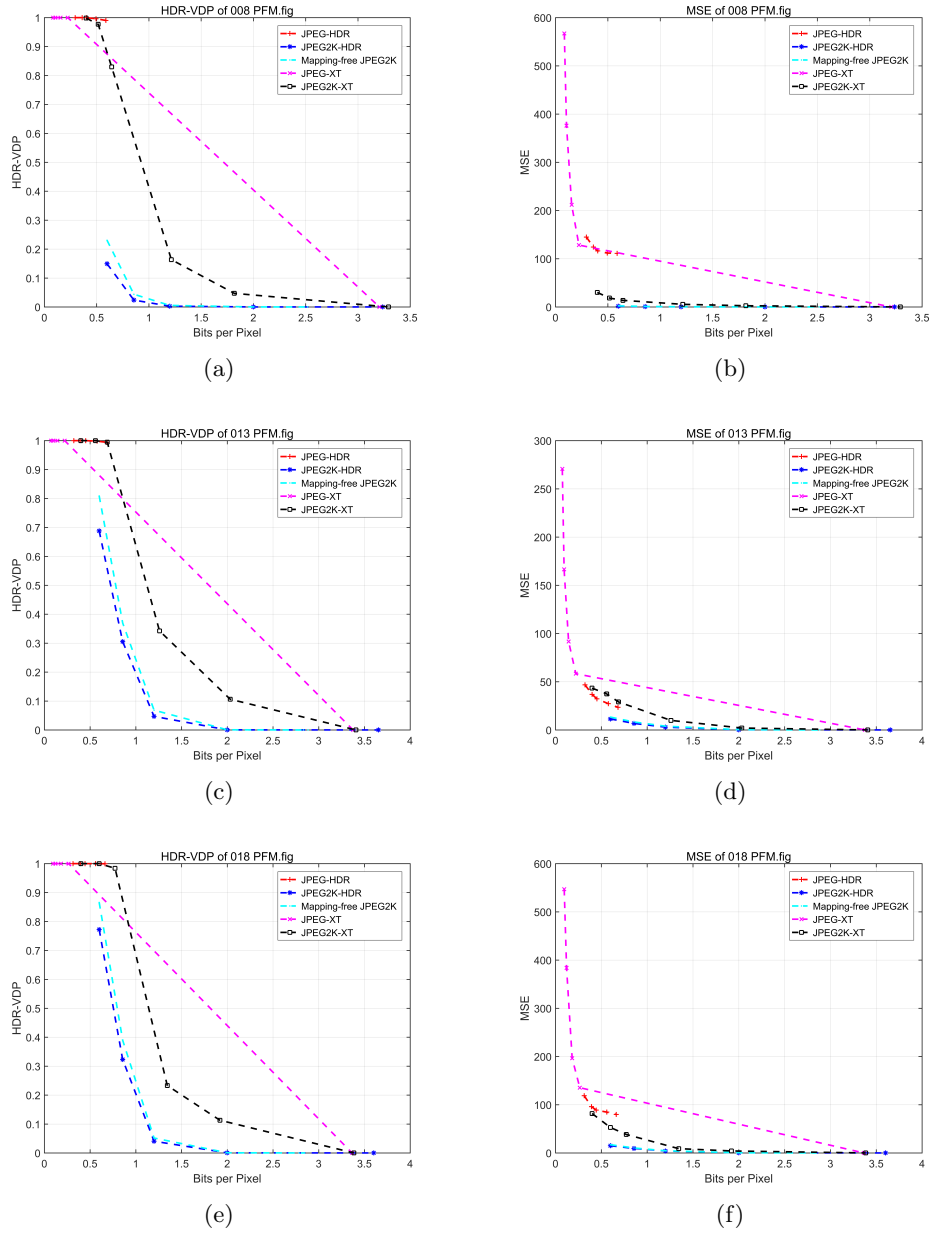


Figure 4.14: Rate-distortion curve of PFM 008/ 013/ 018 images. (a) HDR-VDP result; (b) MSE result; (c)HDR-VDP result; (d)MSE result; (e)HDR-VDP result; (f)MSE result.

Bjontegaard metric describes the distance between two RD curves, which can be seen as a coding gain. The BD-MSE and BD-VDP are used to compute the difference between two corresponding curves.

The BD-MSE and BD-VDP results of images as shown in Table 4.5 represent the relative differences between the proposed mapping-free post HDR

image compression and JPEG-HDR and JPEG2000-HDR. The mapping-free post HDR image compression scheme provides equivalent performance compared to JPEG-HDR and JPEG2000-HDR. The computation complexity is reduced by skipping 'tone-mapping' step. The negative BD values for HDR-VDP or MSE are lower compared to JPEG-HDR or JPEG2000-HDR, which is equivalent to a better quality.

The distortions of JPEG-Xt between quality 99 and 100 increase dramatically. It means for JPEG-Xt encoding, there are no intermediate RD points between the largest two bit-rates. Other JPEG2000 based methods could have smooth RD curve which decreases or increases within 0 and 1 of HDR-VDP. Therefore, the BD results are not very meaningful to compare these compression methods. The full RD results are presented in Appendix C.

The above BD results demonstrated that the mapping-free post HDR image compression method obtained equivalent performance at low bit-rate.

The TMO affect the quantization error because the TMO change the dynamic range of the input content. As a result, the range of transformed coefficients is changed. The quantization errors then increased or decreased by ITMO. The proposed mapping-free method avoids the TMO and ITMO operations, and the quantization error is unaffected. Furthermore, the evaluation illustrated that when HDR points were stored as IEEE floating-point representations, a logarithmic function is already applied on HDR pixels. For the LDR backwards-compatibility, the LDR version can always be acquired by applying TMO to the decoded HDR content. The proposed method can be improved by optimizing the IEEE-754 conversion between integers and HDR floating-points.

The proposed method does not exploit the low-level features of legacy codecs, the compression performances of legacy codecs are not improved. Since the transform, quantization and entropy coding in the encoding are unchanged. These HDR image compression approaches fail to compress non-salient regions differently than the salient regions when HVS perceives them differently, and

Table 4.5: Image Rate Distortion Results

| Image Name | JPEG | JPEG | JPEG | JPEG | JPEG | JPEG |
|--------------------------------|-------|--------------------------|--------|----------------------------|---------|---------|
| | VDP | MSE | 2000 | 2000 | 2000-XT | 2000-XT |
| Apartment.hdr | 0.28 | 0.004 | -0.3 | -9.061 $\times 10^{-4}$ | -0.562 | -0.008 |
| MtTamWest.hdr | 0.526 | 6.64 $\times 10^{-4}$ | -0.633 | -0.005 | -0.886 | -0.013 |
| SpheronSiggraph2001.hdr | 0.028 | 4.608 $\times 10^3$ | -0.057 | -0.005 | -0.003 | 0.237 |
| Flower.exr | NA | NA | 1.193 | -0.002 | 0.162 | 0.001 |
| Ocean.exr | NA | NA | -0.357 | -46.08 | 0.108 | 9.477 |
| StarField.exr | NA | NA | -0.742 | -0.848 | 0.988 | 5.469 |
| McKeesPub.pfm | NA | NA | -0.598 | -275.707 | 0.837 | 109.747 |
| Set22.pfm | NA | NA | -0.43 | -162.017 | 0.393 | 14.412 |
| Set70.pfm | NA | NA | -0.452 | -200.751 | 0.305 | 67.395 |

this is lead to the next Chapter.

4.6.3.3 Video results

Figures 4.15 and 4.16 (a) and (b) show the mean of all frames MSE and HDR-VDP P75 of beerfest_lightshow_01 and bistro_01 sequences. All JPEG2000 based methods out-perform JPEG based method. In JPEG-HDR compression method, the HDR-VDP P75 increased when bit-rate increased.

Figures 4.15 and 4.16 (c) show the HDR-VDP P75 of three Motion JPEG based compression methods for the same sequences. The corresponding bit-rates are: 1.522×10^4 kbps for mapping-free JPEG, $1.462e \times 10^4$ kbps for JPEG-HDR and 1.478×10^4 kbps for JPEG-XT respectively. The blue line is average MSE of mapping-free JPEG method, which is lower than the JPEG-HDR and JPEG-XT methods.

The HDR-VDP P75 is used in [94, 142] to evaluate the HDR video compression quality. This value represents the overall number of pixels which are beyond 75% of detection probability of HDR-VDP, and the lower of HDR-VDP P75 means better compression quality. However, the results suggest that HDR-VDP P75 could not reflect the compression error accurately.

Figures 4.15 and 4.16 (d) describe the MSE of each frame of mapping-

free Motion JPEG for beerfest.lightshow_01 and bistro_01 sequences under different quality settings. The average value of the whole sequence is displayed on the same plot. It can be seen that the distortions of proposed method changed according to the quality setting.

The experimental results illustrate the proposed method obtained equivalent compression performance with less computational complicity by avoiding TMO and ITMO operations.

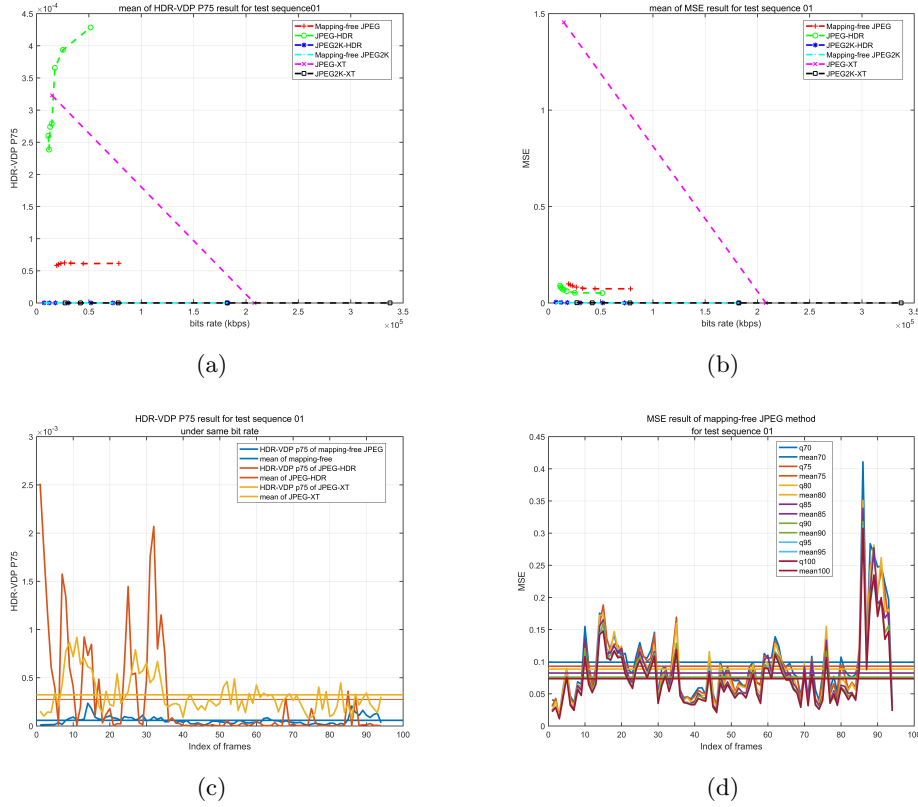


Figure 4.15: Rate-distortion curve of test video sequence 01. (a) HDR-VDP P75 of each frame, (b) MSE of each frame, (c) mean of HDR-VDP P75, (d) mean of MSE. – HDR-VDP P75 represents the overall number of pixels which are beyond 75% of detection probability of HDR-VDP, and the lower of HDR-VDP P75 means better compression quality.

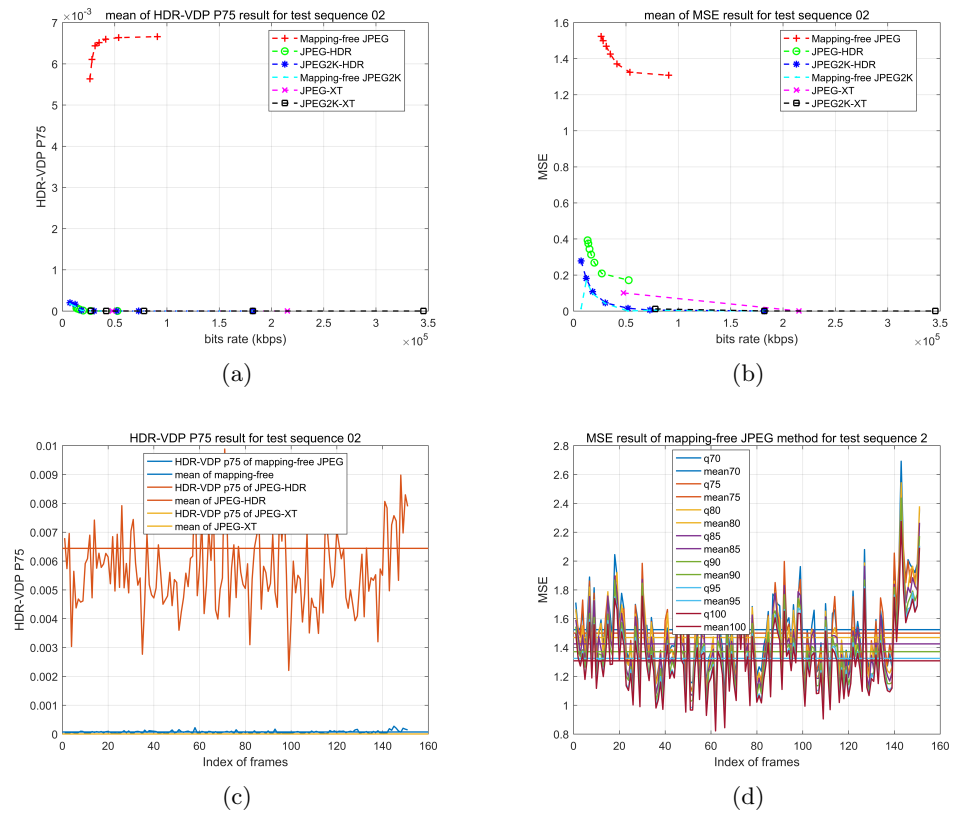


Figure 4.16: Rate-distortion curve of test video sequence 02.
(a) HDR-VDP P75 of each frame, (b) MSE of each frame, (c) mean of HDR-VDP P75, (d) mean of MSE..

HEVC results

The results of HDR mapping-free HEVC encoding are presented in the Figure 4.17. As mentioned in section 4.3.3, the PQ method is no suitable to be compared to the proposed method, because it is not designed for 16-bit encoding and the quality could not increase with extra bit-rate.

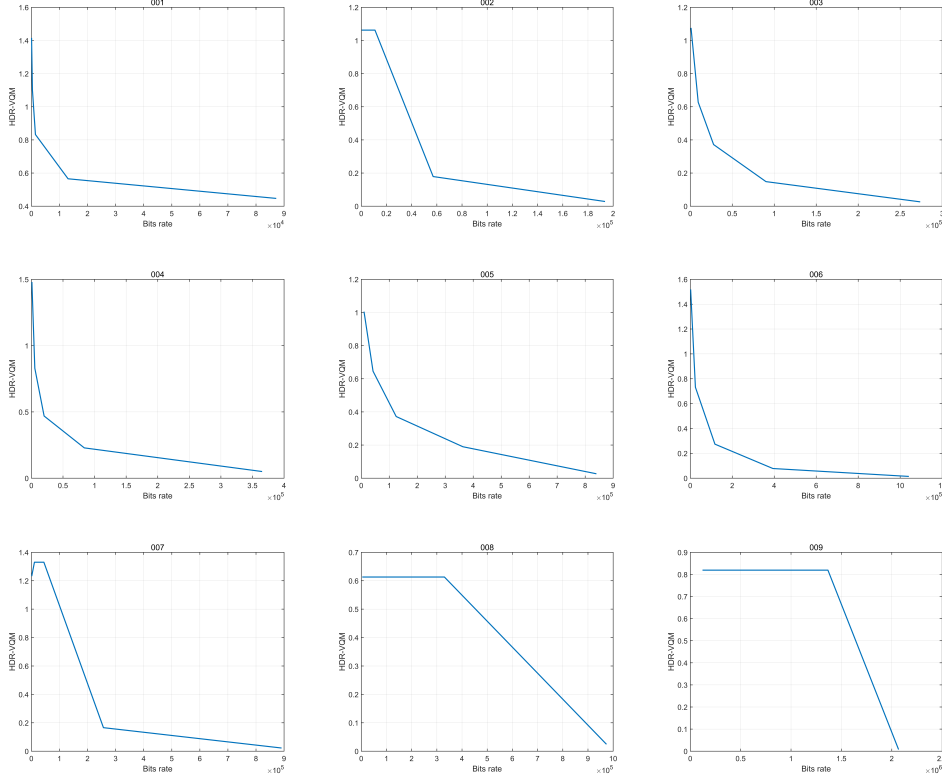


Figure 4.17: Mapping-free HEVC encoding Rate-distortion curve of test video sequence 01-09, from left to right and top to bottom.

4.7 Summary

In this chapter, the importance of the TMO in the HDR compression is evaluated. A mapping-free post HDR image compression based on JPEG and JPEG2000 standard codecs for current HDR image formats, and JPEG2000-XT compression are proposed. The quantization errors affected by TMO and the inverse tone-mapped errors of the quantization errors when decoding the HDR content are eliminated in the proposed method.

The results of mapping-free and JPEG2000-XT HDR image compression are compared to three post HDR image compression methods: JPEG-HDR, JPEG-XT and JPEG2000-HDR. The proposed mapping-free post-HDR image compression scheme has the lowest computational complexity among all methods (50% lower than the previous the state of the art methods) because the tone-mapping or logarithmic transform step is skipped in the proposed method. The compression distortion of the proposed method is comparable to JPEG-HDR and JPEG2000-HDR in MSE and HDR-VDP at low compression ratios. The distortions of JPEG-XT between quality 99 and 100 increase dramatically (from 1 to 0 in HDR-VDP) while the proposed method is able to obtain the intermediate RD points. The proposed mapping-free compression scheme can be extended to video compression.

Chapter 5

Saliency-Guided HDR Visual Content Compression

The shortcomings of the current HDR visual saliency models and HDR visual saliency-guided compression are explored in this chapter. A spatial saliency model for HDR visual content is proposed, more than 90% temporal saliency is predicted by this spatial model. The proposed saliency model can be used to guide HDR compression by applying different quantization factor according to the saliency map.

5.1 Motivation

ROI in various visual content processing applications is determined by saliency detection models. According to [15, 26], at least 70 visual saliency models were proposed for visual saliency prediction, only three of them were proposed for HDR image content. The saliency detection algorithm proposed by Itti *et al.* [58] is the most widely used technique for low dynamic range (LDR) visual attention evaluation. The model predicts bottom-up attention by intensity, colour and orientation features in the Gaussian pyramid, then the saliency maps are computed by centre-surround operations. Itti *et al.* 's model can be extended to HDR image. However, it fails to detect the salient areas of the

median range. The intensity channel in Itti *et al.* 's model has been replaced by the contrast features in CF model since biological sensors are sensitive to contrasts rather than absolute differences. Hence it incorporates the greater dynamic range of HDR content. Gao *et al.* [45, 46] proposed to virtually expose the HDR image to 11 different LDR images and to fuse the Itti *et al.* 's model saliency maps of LDR images. Dong *et al.* [26] suggested to replace Itti *et al.* 's colour channel with a CAM, and luminance channel is converted to units of JND followed by CSF filtering [85].

Because the HDR media records the greater luminance range and wider colour gamut than LDR media, the HDR contents are closer to the actual scene observed by HVS. Thus, saliency detection of HDR content represents more meaningful attention regions of the natural scene. The HDR applications could benefit from the saliency analysis *e.g.* tone-mapping [76] and compression [49].

Lossy HDR image compression JPEG-HDR [131], JPEG2000-HDR [135], JPEG-Xt [3] and Mapping-free HDR image compression [38] have been proposed. These HDR image compression approaches fail to compress non-salient regions differently than the salient regions when HVS perceives them differently. The works of [101, 56, 49] suggested LDR content can be compressed by visual saliency. To extend the LDR visual saliency models for HDR image compression is required in order to compress HDR content by visual saliency.

In this chapter, a photographic visual saliency modeling for HDR images with compression application is proposed. The saliency-guided compression could reduce around 1/3 of bit-rate cost, and it is not essential to compress salient and non-salient regions uniformly. The performance is evaluated by comparing saliency maps, which are computed by computational models and the ground truth from eye tracker based on two quantitative metrics [70]: The PCC and KLD. The results of the evaluation show that the proposed method has the highest average score and the lowest computational complexity among the existing state-of-the-art methods.

5.2 Saliency Models for HDR Image

In order to propose a saliency model for HDR video, the spatial models are investigated first. As will be shown in Section 5.3.4.1, the balance between complexity and accuracy in the current HDR spatial models can be improved. The purpose of the study is to obtain a higher accuracy saliency estimation while keeping the complexity relatively low.

5.2.1 The Proposed Method

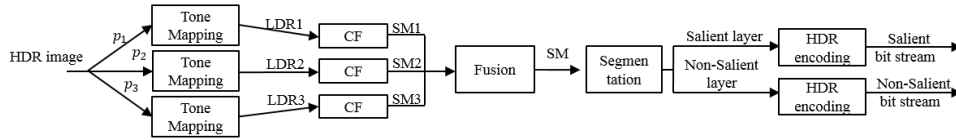


Figure 5.1: The Architecture of the proposed model.

The proposed HDR image visual saliency model is based on the experiment in which the CF is not enough to represent the full dynamic range of HDR content. Furthermore, the 11 of CRF exposed LDR images (Gao) and 9 of CSF filtered adaptation luminance levels (Dong) are redundant to predict the saliency map of whole-dynamic range.

The proposed method is based on the assumption of that the salient regions of a given HDR image after TMO with a stack of three brightness parameters should appear at least once in the saliency maps. Because of the TMO should preserve the HVS perception of the given HDR scene, and for the extreme cases, the extra two 3-f stops over and under exposures should provide the clue for the saliency model.

Figure 5.1 shows the architecture of the proposed model. The proposed HDR saliency map detection method contains 3 parts: tone-mapping, saliency detection, and saliency map fusion. Firstly, the HDR input image is tone-mapped into three LDR images by a stack of brightness parameters of TMO: p_1 , p_2 , p_3 , where p_1 is the default value of TMO, and $p_2 = \frac{1}{8}p_1$, $p_3 = 8p_1$

respectively. Then, the tone mapped LDR images are fed into CF model for generating three saliency maps under different brightness. Finally, the stacks of saliency maps are weighted to obtain the final saliency map of the given HDR image.

The bottom-up visual saliency computational core is based on Itti *et al.*'s saliency model [58] with the contrast feature [17], and the HDR scene is further preserved by the cross dynamic range of three tone-mapped LDR images. The weighting function for LDR image saliency maps fusion is simplified version of [45] with better accuracy. The details of the proposed visual saliency model are described in following sections.

5.2.1.1 TMO

The HDR image is tone-mapped into LDR image by TMO [25, 106]:

$$\begin{aligned} I_{l,1} &= \mathcal{T}_{p_1}(I_h), \\ I_{l,2} &= \mathcal{T}_{p_2}(I_h), \\ I_{l,3} &= \mathcal{T}_{p_3}(I_h), \end{aligned} \tag{5.1}$$

where I_l and I_h are LDR and HDR image respectively, T can be any kind of TMOs, p_1, p_2, p_3 are brightness parameters of TMO, p_1 is the default value of TMO, and $p_2 = \frac{1}{8}p_1$, $p_3 = 8p_1$ respectively.

The number three is chosen for *two* reasons. Firstly, less LDR images are equivalent to less complexity of the model. Secondly, single LDR or HDR input is not enough to analyse HDR visual attention regions completely. In the proposed method, the first LDR image is created by the default TMO parameters p , and the other two LDR images are the 8 and the 1/8 brighter and darker versions. These three images ensure the coverage of 6 f-stops dynamic range changes connectedly, and each LDR image contains around 8 f-stops itself.

5.2.1.2 Saliency Detection

The bottom-up visual saliency computational core is based on Itti *et al.* saliency model [58] with improved contrast feature [17] for both luminance and orientation channels from equation (5.2) and (5.4) to (5.3) and (5.5).

$$I'(c, s) = |I(c) \ominus I(s)|, \quad (5.2)$$

$$I''(c, s) = \frac{|I(c) \ominus I(s)|}{I(s)}, \quad (5.3)$$

where I and O are luminance and orientation channels of Itti *et al.*'s model, and \ominus is the center-surround operator, the across-scale difference obtained by interpolating the coarser scale to the finer scale and then subtracting them pixel by pixel. The c and s are scales in the Gaussian pyramids.

$$O'(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)|, \quad (5.4)$$

$$O''(c, s, \theta) = \frac{O(c, \theta)}{I(s)}, \quad (5.5)$$

Hence, the saliency maps of three LDR images can be obtained by CF model:

$$\begin{aligned} S_{l,1} &= C(I_{l,1}), \\ S_{l,2} &= C(I_{l,2}), \\ S_{l,3} &= C(I_{l,3}), \end{aligned} \quad (5.6)$$

where S is the saliency map, I_l is LDR image, C is the CF model. Figure 5.2 shows the example of the given HDR images and their tone mapped LDR version and CF saliency map results.

5.2.1.3 Saliency Fusion

The proposed model tends to exploit visual saliency regions which is fully from three LDR image. In [45][46], the most of weights are given to two

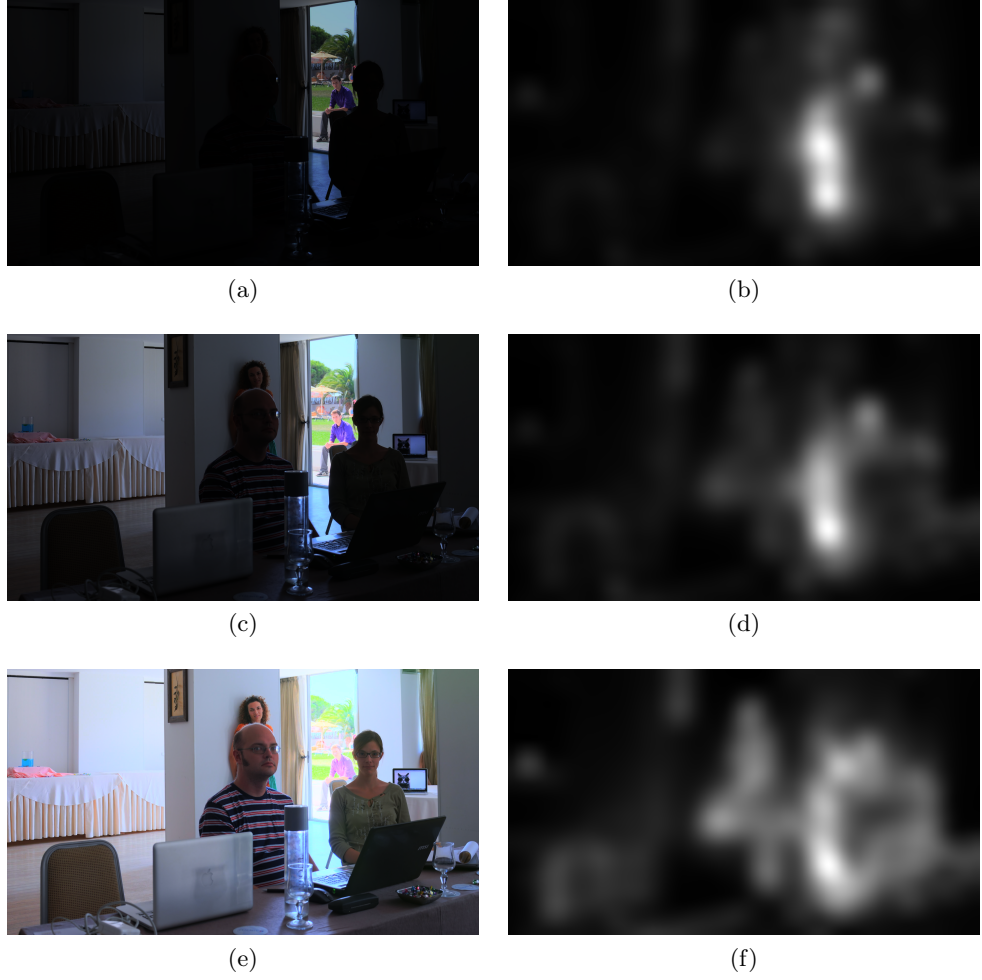


Figure 5.2: (a,b) Tone mapped under exposure image and corresponding saliency map, (c,d) Tone mapped proper exposure image and corresponding saliency map, (e,f) Tone mapped over exposure image and corresponding saliency maps.

neighbouring weights which are the nearest pair to the 'perfect' exposure time version, the weights of others' saliency maps are negligible according Gao *et al.* weighting function, see equation (5.7).

$$w(\Delta t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left(\log_2 \frac{\Delta t}{\Delta \bar{t}}\right)^2\right), \quad (5.7)$$

For example:

for $\Delta t = [1/15, 1/8, 1/4, 1/2, 1, 2, 4, 8, 15]$,

the corresponding $w = [0.005, 0.036, 0.147, 0.301, 0.308,$

0.158, 0.040, 0.005, 0.0004]. The fusion step of the proposed method is finished by taking the highest saliency value for each pixel location among three saliency maps see equation (5.8), which is also accomplished by the assumption made:

$$SM(x, y) = \max(S_{(1,2,3)}(x, y)); \quad (5.8)$$

5.3 Saliency Model for HDR Video

To complete the whole saliency model for the HDR image, the temporal saliency needs to be predicted. Since the literature suggests that [80] the temporal saliency is over-weight spatial saliency in the video due to HVS perception. The following subsections demonstrate the performance evaluation of the existing temporal saliency models.

5.3.1 LDR Temporal Models

The existing LDR temporal saliency estimation methods operate on the temporal domain and calculate the difference between frames and motion vectors. These models work on the same principle and they achieve closed performance.

At the beginning of the investigation, there is no HDR video saliency ground truth available. Only few LDR ground truth of visual saliency datasets are available ^{1 2 3} [124, 43, 48].

However, their contents have similar disadvantages:

- low resolution (CIF 352 by 288 pixels)
- binary classified
- usually single object defined
- no salient regions appear in the background

Hereby, some examples for LDR visual saliency dataset are listed below:

¹<http://cpl.cc.gatech.edu/projects/SegTrack/>

²<http://www.kecl.ntt.co.jp/people/kimura.akisato/saliency3.html>

³<http://www.cc.gatech.edu/cpl/projects/videosegmentation/>

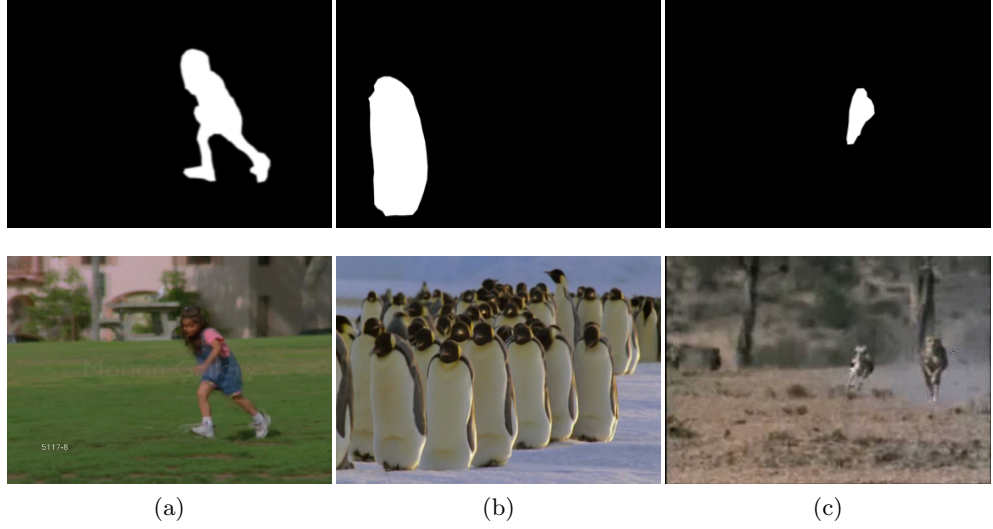


Figure 5.3: Example of data set,(a) girl (b) penguin,(c) cheetah

The sequences penguin and cheetah are not used, because the saliency objectives are not well defined in the ground-truth. (see Figure 5.3) The existing classic temporal saliency models are tested on a few sequences, and they represent different approaches to measure the motion differences between two neighbouring frames, by calculating the absolute differences, motion vector, optical flow, and differences in the frequency domain between two consecutive frames respectively:

Table 5.1: Pearson Correlation of different methods

| Sequence | IttiF[57] | IttiM[57] | Dong[28] | Oakes[98] |
|-----------|--------------|-----------|----------|-------------|
| girl | 0.198 | 0.061 | 0.034 | 0.41 |
| parachute | 0.603 | 0.272 | 0.178 | 0.43 |
| birdfall2 | 0.585 | 0.564 | 0.48 | 0.04 |

Where IttiF and IttiM are the flicker and motion temporal saliency maps of [57]. The higher correlation of temporal estimation map means more accurate predictions. The results in Table 5.1 suggest there is no particular method is optimal in all LDR scenarios. The prediction accuracies of these methods are close.

5.3.2 HDR Temporal Models

To the best knowledge of the author of this thesis, there is no existing temporal saliency model for predicting HDR contents. Then the performances of these temporal models on the HDR contents is evaluated. The current LDR temporal saliency models use the relative differences between two consecutive frames to describe the temporal changes. The major difference between the HDR and the LDR contents is the dynamic range, and the relative differences between frames will almost always apply the normalization procedure to acquire the temporal saliency map. Therefore, the normalization procedure is dynamic range independent, and LDR temporal model can be applied to the HDR sequence.

The experimental results suggest that the most of the temporal saliency maps predicted by the classic models belong to the spatial saliency maps to a large extent. To prove this assumption, a measurement which only quantizes the temporal saliency predicted by temporal model is defined as:

- Calculate both spatial saliency S_m and temporal saliency T_m maps for video.

- There are overlaps between spatial saliency and temporal saliency maps.

$$S_{(i,j)} \cap T_{(i,j)}$$

- Get the indices of unpredicted and under-predicted temporal saliency regions by comparing the temporal saliency map to the spatial saliency map.

$$S_{(i,j)} < T_{(i,j)}$$

- Calculate the total volume D of un-predicted and under-predicted temporal saliency regions.

The volumes of saliency map is the intensity summation of the whole saliency map. If considering the saliency map as a three dimensional matrix where the intensity is the Z axis.

The unpredicted and under-predicted temporal saliency regions are the regions with lower intensity in the spatial saliency map, compared to the temporal map. The total volume D of unpredicted and under-predicted temporal saliency regions is given by:

$$D_{(i,j)} = \begin{cases} \sum(T_{(i,j)} - S_{(i,j)}), & \text{if } (T_{(i,j)} - S_{(i,j)}) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (5.9)$$

where i, j are the pixel indices.

Table 5.2: Differences between S_m and T_m
(The thumbnails of these sequence can be seen in Appendix A Figure A6)

| Sequence | $\sum S$ | $\sum T$ | D |
|------------------------------|--------------------|--------------------|--------------------|
| beerfest lightshow 06 | | | |
| IttiF | 1.37×10^6 | 3.7×10^5 | 2.12×10^4 |
| IttiM | / | 1.73×10^5 | 2.79×10^4 |
| Oakes | / | 1.74×10^5 | 3.8×10^3 |
| Dong | / | 5.6×10^4 | 3.33×10^3 |
| bistro 01 | | | |
| IttiF | 1.25×10^8 | 2.82×10^7 | 5.53×10^6 |
| IttiM | / | 2.69E+07 | 3.99×10^6 |
| Oakes | / | 3.46×10^5 | 2.71×10^5 |
| Dong | / | 1.78×10^7 | 1.95×10^6 |
| carousel fireworks 01 | | | |
| IttiF | 1.42×10^8 | 9.75×10^6 | 9.41×10^5 |
| IttiM | / | 1.08×10^7 | 1.25×10^6 |
| Oakes | / | 9.17×10^6 | 1.15×10^5 |
| Dong | / | 3.48×10^6 | 1.07×10^5 |
| cars fullshot | | | |
| IttiF | 1.39×10^8 | 1.20×10^7 | 1.40×10^6 |
| IttiM | / | 1.30×10^7 | 1.87×10^6 |
| Oakes | / | 7.30×10^6 | 1.71×10^5 |
| Dong | / | 1.32×10^7 | 1.84×10^6 |
| fireplace 01 | | | |
| IttiF | 1.06×10^6 | 2.3×10^5 | $\times 10^4$ |
| IttiM | / | 2×10^5 | 1.08×10^4 |
| Oakes | / | 4.06×10^4 | 354.18 |
| Dong | / | 2.02×10^5 | 6.76×10^4 |
| fishing longshot | | | |
| IttiF | 1.25×10^6 | 4.23×10^5 | 2.52×10^4 |
| IttiM | / | 2.53×10^5 | 3.1×10^3 |
| Oakes | / | 8.19×10^5 | 1.91×10^5 |
| Dong | / | 2.29×10^5 | 2.46×10^4 |
| poker fullshot | | | |
| IttiF | 1.89×10^6 | 8.39×10^4 | 1.32×10^4 |
| IttiM | / | 1.28×10^5 | 1.37×10^4 |
| Oakes | / | 1.22×10^5 | 5.05×10^3 |
| Dong | / | 4.61×10^4 | 641.32 |
| showgirl 01 | | | |
| IttiF | 1.61×10^6 | 1.05×10^5 | 576.04 |
| IttiM | / | 1.33×10^5 | 2.24×10^3 |
| Oakes | / | 3.46×10^5 | 172.71 |
| Dong | / | 6.69×10^4 | 4.22×10^3 |
| smith hammering | | | |
| IttiF | 1.38×10^6 | 1.75×10^5 | 756.6 |
| IttiM | / | 1.66×10^5 | 841.75 |
| Oakes | / | 6.39×10^4 | 263.83 |
| Dong | / | 3.59×10^4 | 3.19×10^3 |

The theoretical maximum volume of saliency map S is $1920 * 1080 * 255 = 5.29E + 08$.

As shown in Table 5.2, the unpredicted and under-predicted temporal regions (which is equal to $D/\sum T$) are relatively low – about 7% average of these sequences. An example of a sequence is given in Figure 5.4, The D is much lower than the T and S . This finding means that the around 93% of the temporal saliency information is already predicted in the spatial saliency model. Therefore, the temporal saliency prediction and the spatial-temporal fusion are not always necessary. Figure 5.5 to 5.8 show examples of this finding.

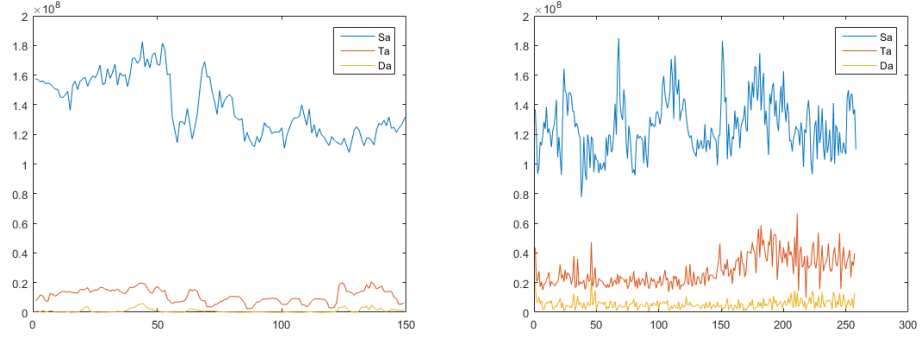


Figure 5.4: Examples of a set of S_m , T_m , D_m .

The HDR video saliency ground truth is available after [27] fixed their server. Therefore the temporal correlation can be evaluated by using the subjective evaluation results. As illustrated by the results in Table 5.3, there is no particular method is optimal in all HDR scenario, and the additional temporal saliency information predicted by the temporal model is lower than 10%. Hence, the proposed model can have two options, the spatial model alone for most of the cases and spatial-temporal model for the extreme high motion cases. It is worth mentioning that subjective experiment [27] was conducted at a lower resolution (1024 by 576 pixels) compared to native visual content (1920 by 1080 pixels).

Table 5.3: Correlation between temporal saliency methods and the ground truth fixation density map

| Seq | bistro1 | bistro2 | bistro3 | carousel1 | carousel8 | fishing |
|-------|--------------|--------------|--------------|--------------|--------------|--------------|
| IttiF | 0.039 | 0.114 | 0.06 | 0.286 | 0.124 | 0.031 |
| IttiM | 0.059 | 0.025 | 0.037 | 0.084 | 0.074 | 0.033 |
| Oakes | 0.096 | 0.154 | -0.009 | 0.486 | 0.076 | 0.128 |
| Dong | 0.134 | 0.116 | 0.138 | -0.028 | 0.129 | 0.073 |



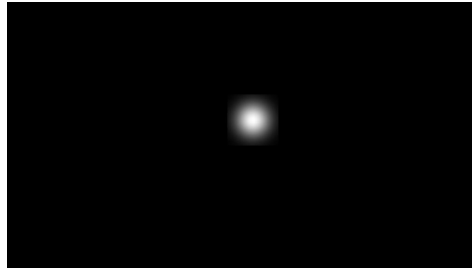
(a)



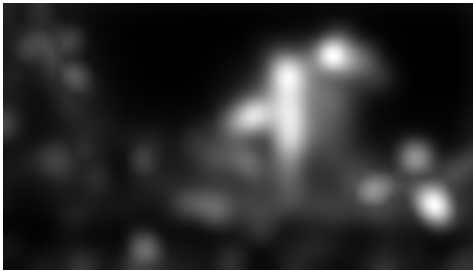
(b)



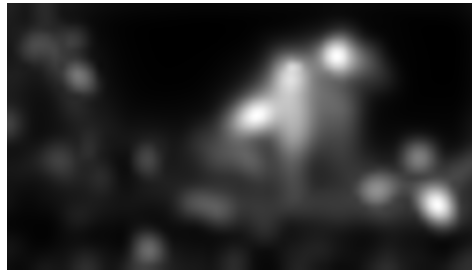
(c)



(d)



(e)



(f)

Figure 5.5: (a,b) The consecutive frames (1 and 2), (c,d) The ground truth, (e,f) The saliency map of the proposed spatial model.

5.3.3 Spatial-Temporal Model Fusion

As can be seen in the Table.5.6 and Table.5.7, the proposed spatial model is more accurate than other spatial methods. If the proposed spatial model follows the same fusion rule with the same temporal saliency predictions, the

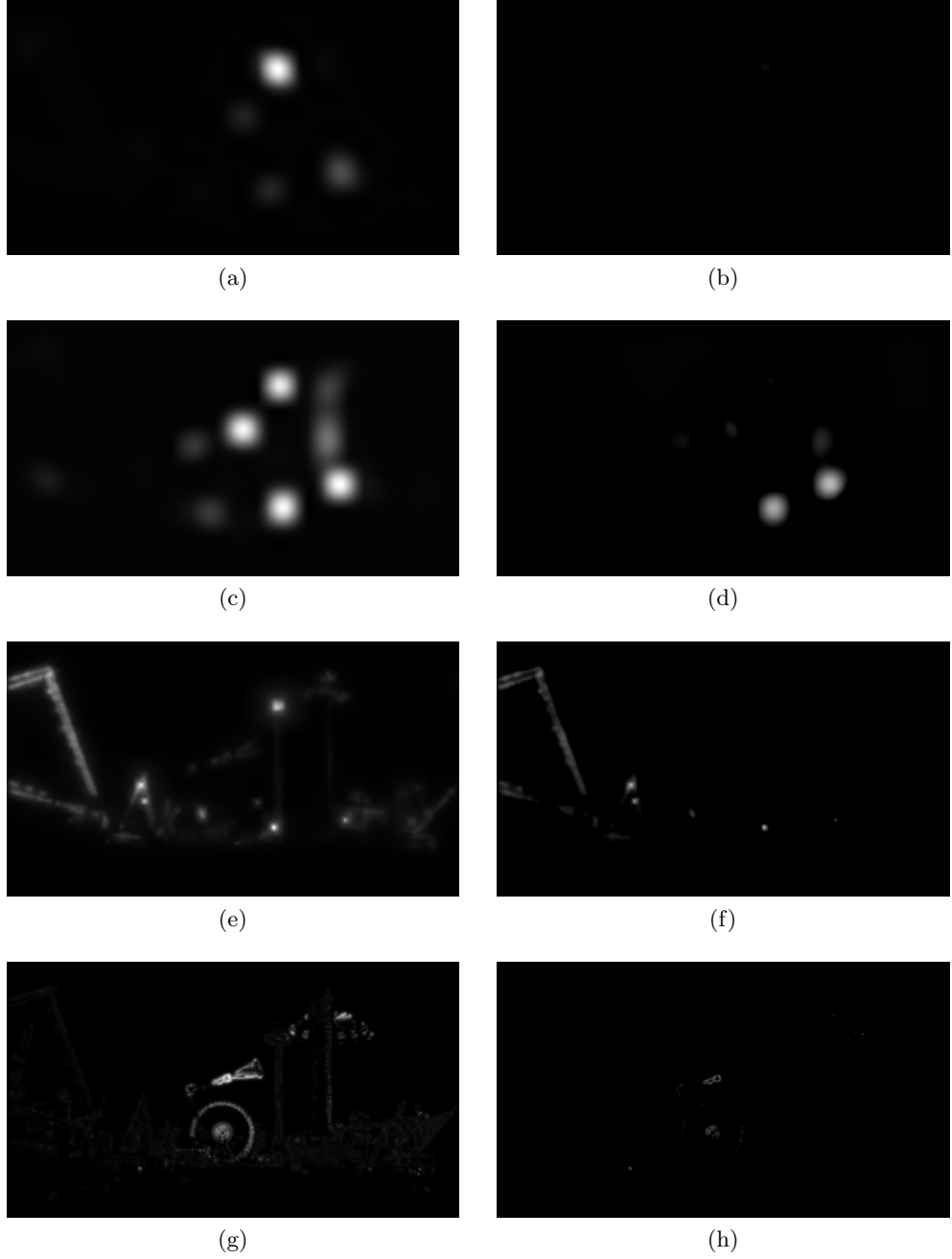


Figure 5.6: (a,b) The temporal saliency map of IttiM model and the unpredicated information by the proposed spatial model, (c,d) The temporal saliency map of IttiF model and the unpredicated information by the proposed spatial model, (e,f) The temporal saliency map of Oaks model and the unpredicated information by the proposed spatial model, (g,h) The temporal saliency map of Dong model and the unpredicated information by the proposed spatial model.

result of the proposed spatial-temporal model should achieve a higher accuracy than existing methods. As given in the previous section temporal saliency

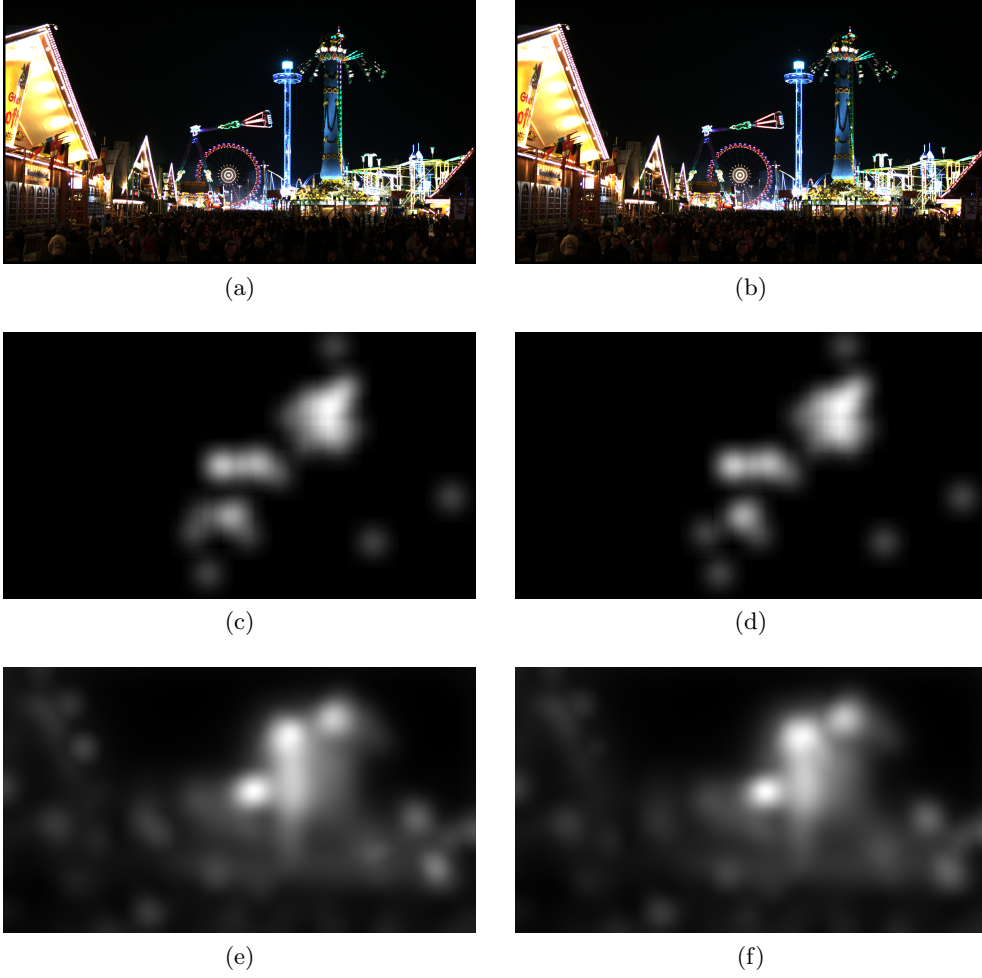


Figure 5.7: (a,b) The consecutive frames (30 and 31), (c,d) The ground truth, (e,f) The saliency map of the proposed spatial model.

is not used as a guidance for compression scheme, because around 93% of the temporal saliency information is already predicted in the spatial saliency model. This subsection describes the spatial-temporal rule in case that other studies prove otherwise.

The spatio-temporal saliency map is given by:

$$S = a \cdot S(s) + b \cdot S(t) + c \cdot S(s) \cdot S(t), \quad (5.10)$$

where a , b and c are weighting factors of spatial, temporal and spatial-temporal cross channel saliency maps [80].

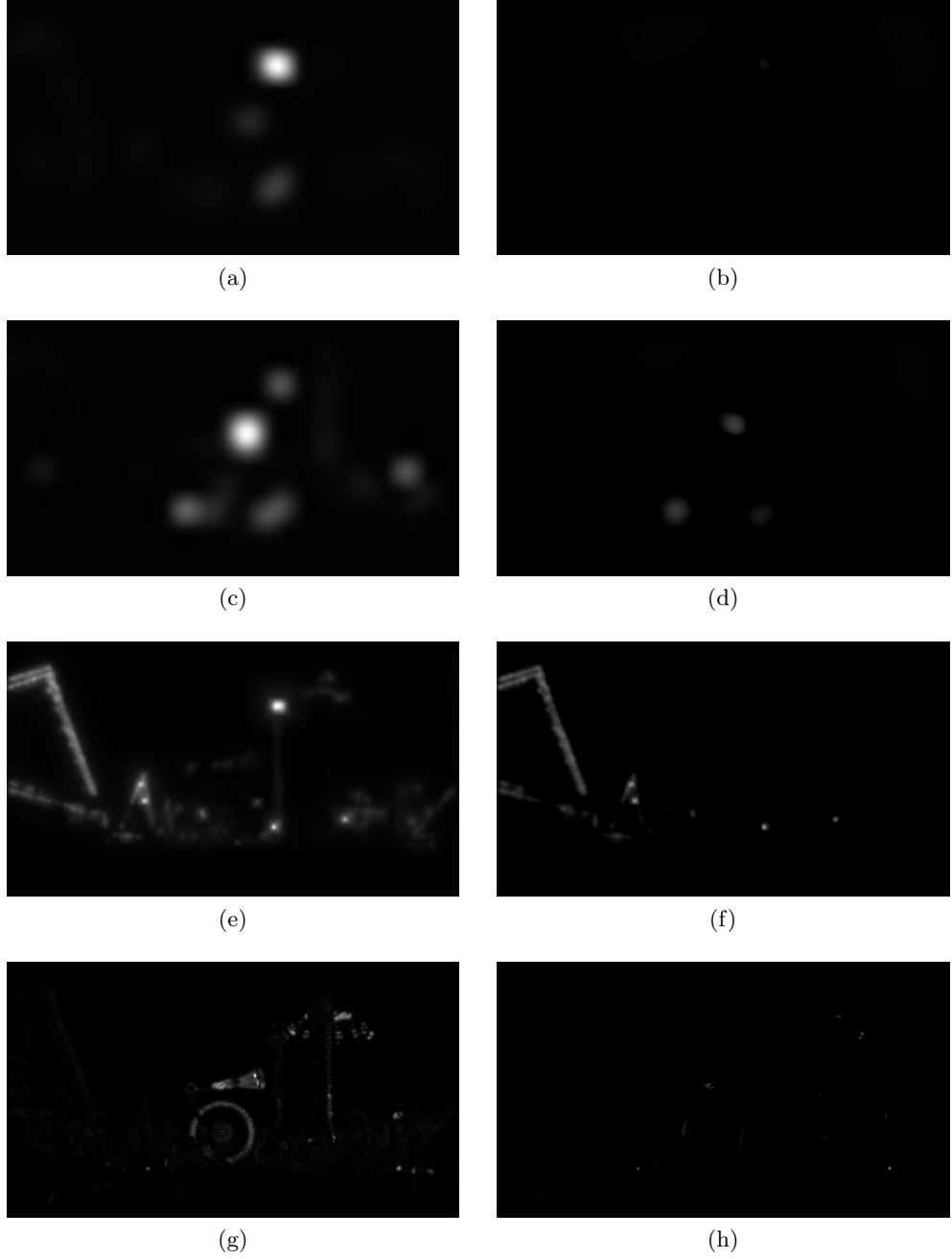


Figure 5.8: (a,b) The temporal saliency map of IttiM model and the unpredicated information by the proposed spatial model, (c,d) The temporal saliency map of IttiF model and the unpredicated information by the proposed spatial model, (e,f) The temporal saliency map of Oaks model and the unpredicated information by the proposed spatial model, (g,h) The temporal saliency map of Dong model and the unpredicated information by the proposed spatial model.

5.3.4 Experimental Results and Discussion

In this section, the results of the proposed spatial saliency model are evaluated.

The ROC is not used as a metric to evaluate the accuracy of proposed method because both prediction maps and ground truth maps are density map, hence the different thresholds of binalization affects the ROC result, as shown in Figure 5.9.

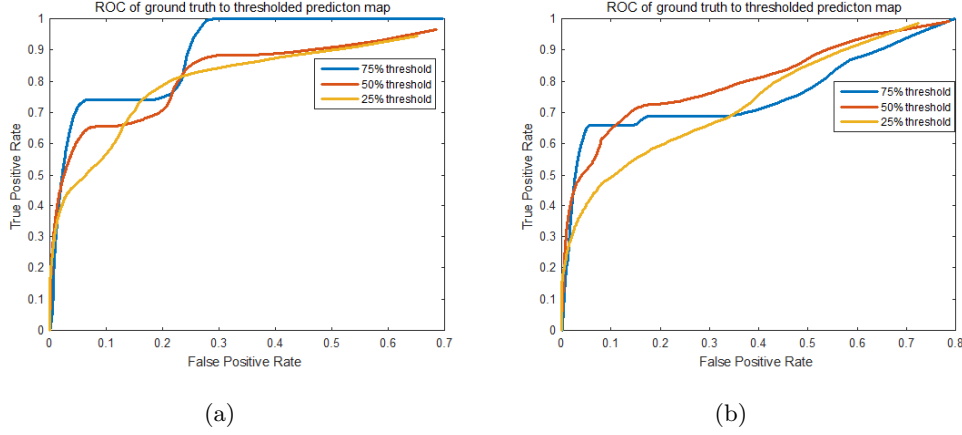


Figure 5.9: The ROC under different thresholds of binalization of No.14 image and right for No.33 image.

5.3.4.1 Computational Complexities evaluation

The computational complexity of the state of the art HDR spatial saliency models and the proposed method are shown in Table 5.4. The assumed complexity of Itti *et al.* 's model is $I(n)$, and all HDR models are based on this model. The complexity of CF model is approximately the same to Itti *et al.* 's model. The Gao *et al.* 's model computes Itti *et al.* 's saliency map on each CRF virtual exposed image, and there are 11 exposures. For Dong *et al.* 's model, the intensity channel is first converted to luma of the JND space $nO(M(1)k)$ and then the luma pass through the multi-CSF in the FFT domain, with a equivalent complexity of $7nO(n \log(n))$. The RGB to LMS colour space conversion and subtraction results in the complexity of $5nO(M(1)k)$.

The proposed method has lower computational complexity than Gao *et al.* 's model and Dong *et al.* 's model and higher complexity than Itti *et al.* 's model and CF *et al.* 's model.

Table 5.4: Computational Complexities (per frame)

| Spatial Saliency Model | Complexities |
|-------------------------------|--------------------------------------|
| Itti | $I(n)$ |
| CF | $I(n)$ |
| Gao | $11nO(M(1)k) + 11I(n)$ |
| Dong | $7nO(n \log(n)) + 5nO(M(1)k) + I(n)$ |
| Proposed | $3nO(M(1)k) + 3I(n)$ |

The actual computational time is computed by a i5-4690K 3.5GHz CPU, 8GB RAM, MATLAB 2016b. Where on the average TMO takes 0.32 secs. The Itti's model takes 0.12 secs on the average – when the width of the output saliency map is 64 pixels. The real computational time of these method are displayed in Table5.5.

Table 5.5: Computational time (secs per frame)

| Spatial Saliency Model | Computational time |
|-------------------------------|---------------------------|
| Itti | 0.12 |
| CF | 0.12 |
| Gao | 4.84 |
| Dong | 2.34 |
| Proposed | 1.32 |

5.3.4.2 Image Results

In the Table.5.6 and Table.5.7, the PCC and KLD results of the proposed method have the highest mean values for PCC (the higher the better) and lower KLD (the lower the better) against the existing state-of-the-art methods, which means the proposed method has the closest predictions to the ground truth from eye tracking data. However, the standard deviation of proposed method is not the lowest, this is due to the proposed method boost the overall contrast and in some cases, not all high local contrast regions are salient in the subjective evaluation.

There are other LDR saliency estimation models that can be extend to HDR model by adopting CF. However, the results are not as accurate as HDR models.

Table 5.6: The result of the Pearson Correlation.

| no. | Itti | CF | Gao | Dong | Proposed |
|-------------|-------------|------|-------------|-------------|-------------|
| 14 | 0.12 | 0.26 | 0.41 | 0.4 | 0.66 |
| 15 | -0.02 | 0.15 | 0.46 | 0.15 | 0.3 |
| 17 | 0.80 | 0.73 | 0.26 | 0.62 | 0.53 |
| 20 | 0.15 | 0.17 | 0.36 | 0.34 | 0.16 |
| 22 | 0.12 | 0.14 | 0.19 | 0.03 | 0.08 |
| 24 | 0.41 | 0.39 | 0.30 | 0.44 | 0.47 |
| 33 | 0.07 | 0.39 | 0.20 | 0.24 | 0.62 |
| 39 | 0.42 | 0.38 | -0.02 | 0.52 | 0.26 |
| Mean | 0.26 | 0.33 | 0.27 | 0.34 | 0.38 |
| Std | 0.27 | 0.20 | 0.15 | 0.22 | 0.21 |

Table 5.7: The result of the absolute value of Kullback-Leibler Divergence.

| no. | Itti | CF | Gao | Dong | Proposed |
|-------------|-------------|-------------|-------------|------|-------------|
| 14 | 0.42 | 0.19 | 0.35 | 1.61 | 0.44 |
| 15 | 0.24 | 0.68 | 0.28 | 1.21 | 0.41 |
| 17 | 1.16 | 0.48 | 0.93 | 1.57 | 0.38 |
| 20 | 0.54 | 0.91 | 0.72 | 1.24 | 0.75 |
| 22 | 0.25 | 0.62 | 0.38 | 0.61 | 0.22 |
| 24 | 0.67 | 0.64 | 0.86 | 0.9 | 0.21 |
| 33 | 0.39 | 0.99 | 0.34 | 1.26 | 0.61 |
| 39 | 0.65 | 0.60 | 0.70 | 0.77 | 0.4 |
| Mean | 0.54 | 0.64 | 0.57 | 1.14 | 0.42 |
| Std | 0.30 | 0.25 | 0.26 | 0.36 | 0.18 |

In the Figure 5.10, the luminance channel in Itti *et al.* 's model is replaced by the CF with normalization. The results suggest the DWT decomposition saliency model is sensitive to median-frequency components which are not always the true salient regions defined by the ground truth. Although it can be improved by replacing luminance channel with contrast channel f to i , l to k . However, because other features are ignored in this method, it has lower accuracy than the other methods.

The first rows of Figures 5.11, 5.13, 5.15, 5.17 show tone-mapped HDR images of dataset, the remaining rows of Figures 5.11 to 5.3.4.2 represent the corresponding saliency maps predicted by different models and the ground truth as shown in the second rows of Figures.



Figure 5.10: (a) HDR image; (b) luminance; (c) ground truth, (d) Weber's contrast, (e) normalized contrast, (f) DWT model of HDR, (g) DWT Weber's contrast model of LDR, (h) DWT normalized contrast model of LDR, (i) DWT model of LDR, (j) Dong's model, (k) CF model, (l) Itti's model of HDR, (m) proposed model.

In Figure 5.11, the blue-ball in the middle-left of the image no.14 is detected as a salient object by Gao's model and the proposed model because others methods are less sensitive to low luminance area.

In Figure 5.17, the window at the center of the image no.33 is salient in

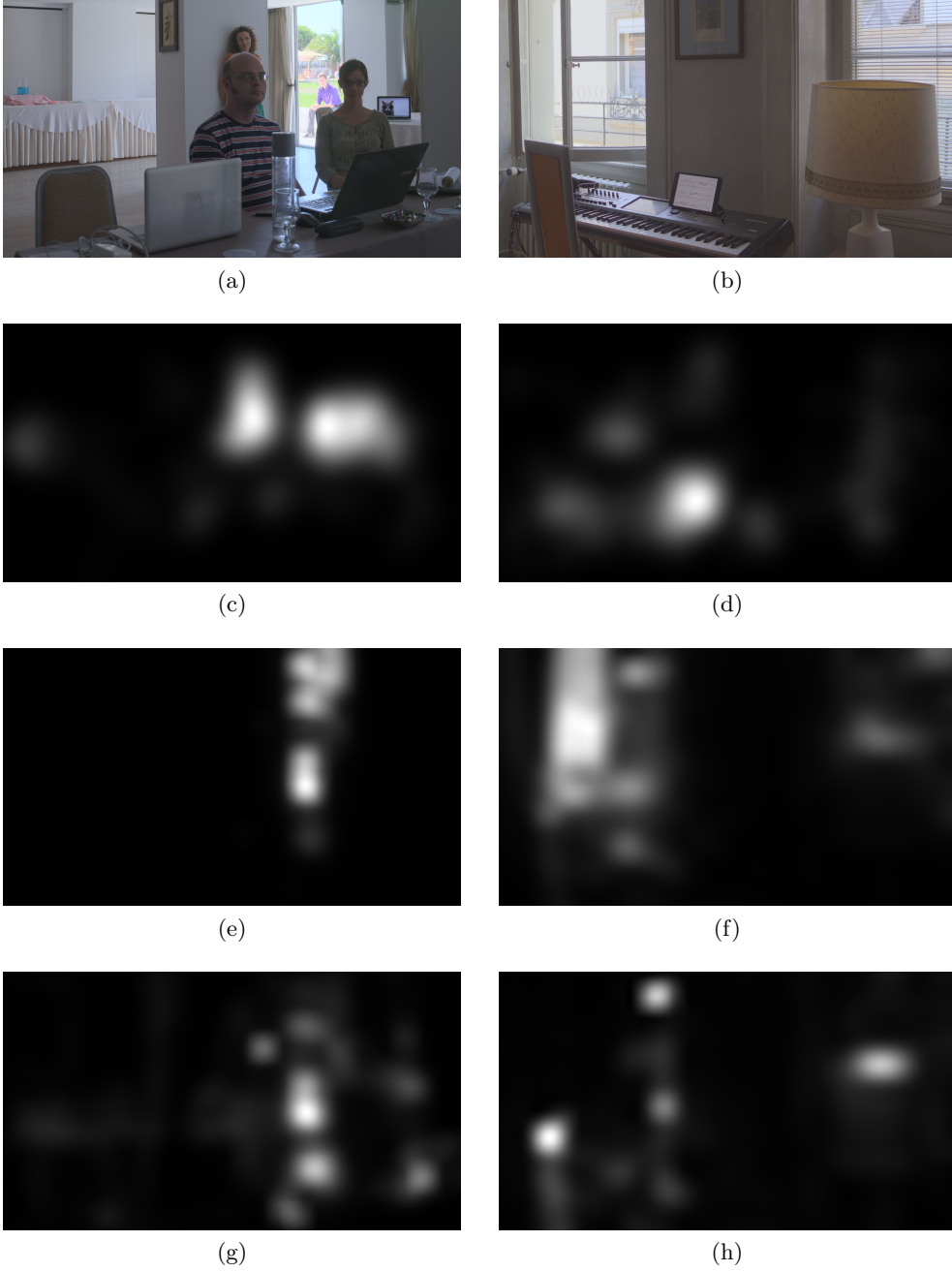


Figure 5.11: Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 14 image and right for No 15 image.

the ground truth and the proposed method while other predictions failed to detect it.

Overall, the proposed method is able to detect the most objective compared to these counterparts, because of the TMO and the stack of tone-mapped LDR

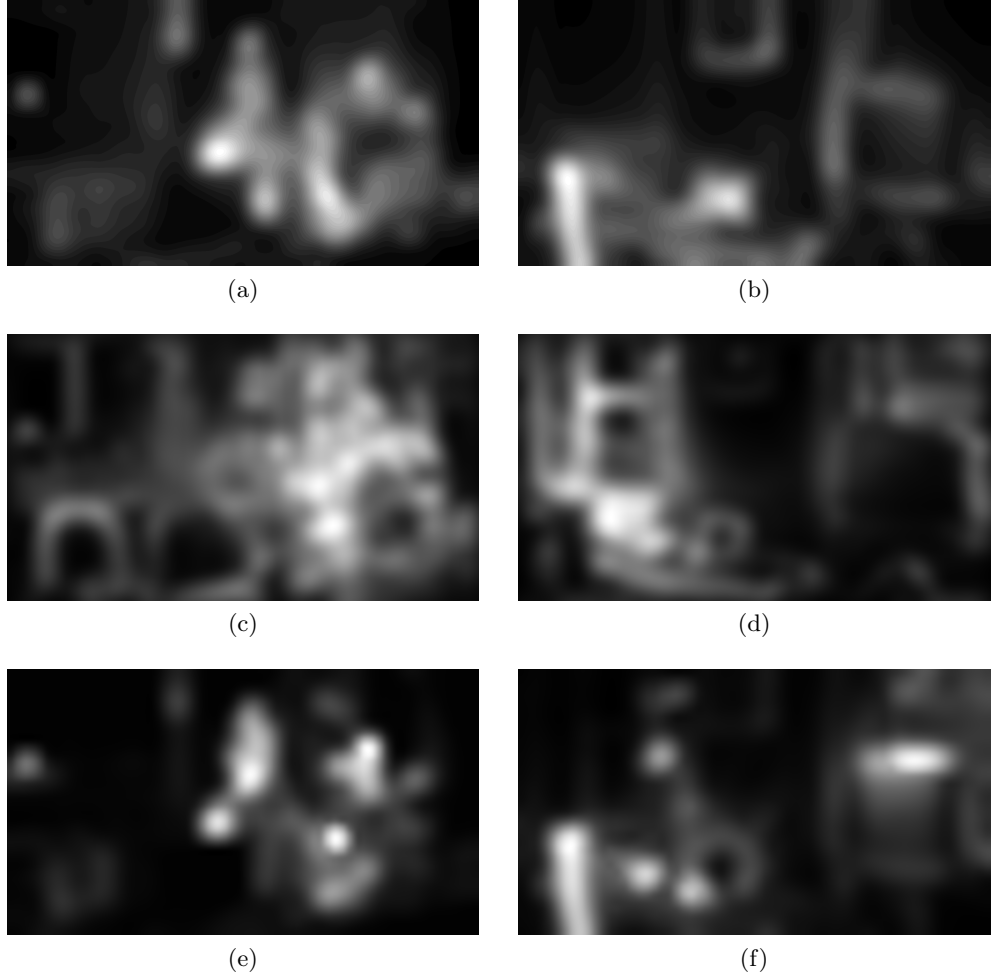


Figure 5.12: (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 14 image and right for No 15 image.

images preserved more local and global contrast compared to other methods.

note: Since there is no guaranteed peak value in Gao's weighting function see equation (5.7), in some cases, there are discontinuous artifacts for normalization.

5.3.4.3 Discussion

The existing LDR temporal visual saliency models operated similarly, they measuring the differences between two consecutive frames by using different approaches. These models can be extended to HDR sequence, by defining a relationship between the spatial saliency map and the temporal saliency

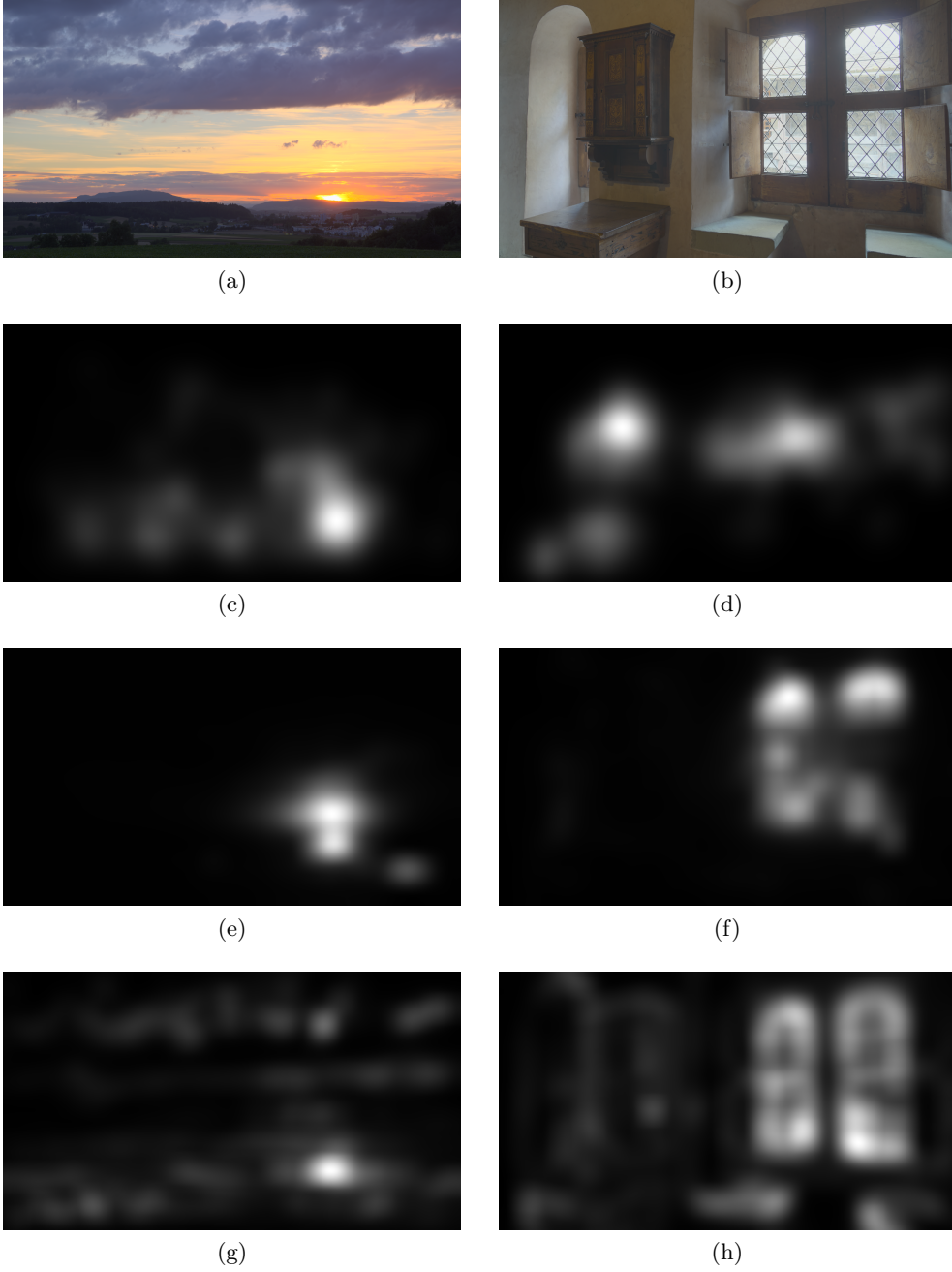


Figure 5.13: Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 17 image and right for No 20 image.

map, it is concluded that more than 93% on the average of temporal saliency are already predicted by the spatial saliency model, for both LDR and HDR sequence.

The proposed spatial visual saliency adopted cross features to provide

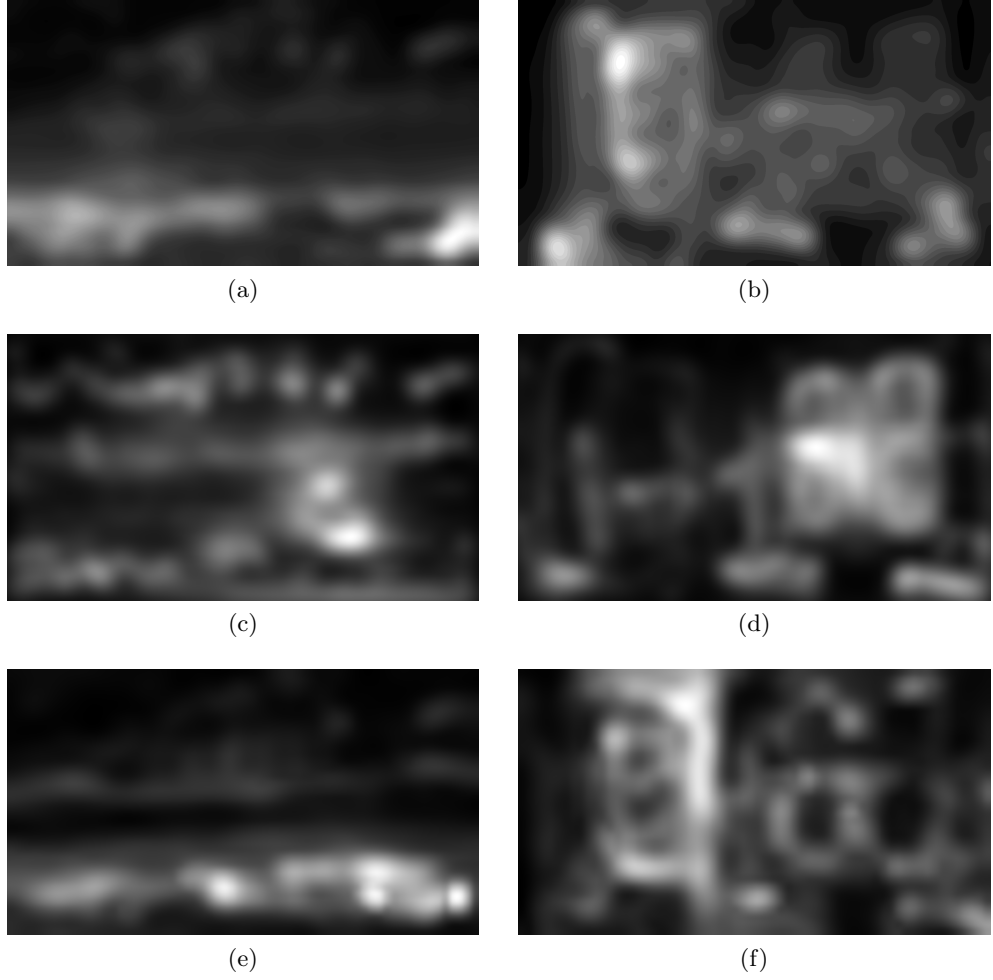


Figure 5.14: (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 17 image and right for No 20 image.

wider dynamic range detection ability while maintaining low computational complexity. The proposed method is able to detect every possible salient object because of the TMO, CF and weighting rules are more sensitive than other methods. However, some salient objects are not the true salient regions in the ground truth.

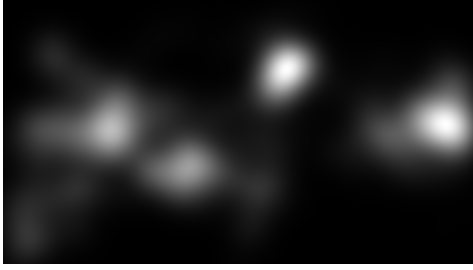
The evaluation is based on the only available HDR video visual saliency ground truth, and the results suggest that the more than 90% of temporal saliency already predicted by the spatial model, this is also has been shown on the evaluation of LDR content. This can be interpreted as the temporal saliency model is not necessary for compression application, the computational



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)

Figure 5.15: Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 22 image and right for No 24 image.

power can be saved. Because of the HDR acquisition limitations, there is no high-motion sequence available in HDR format, the relative motion intensities of the sequences are around the same level.

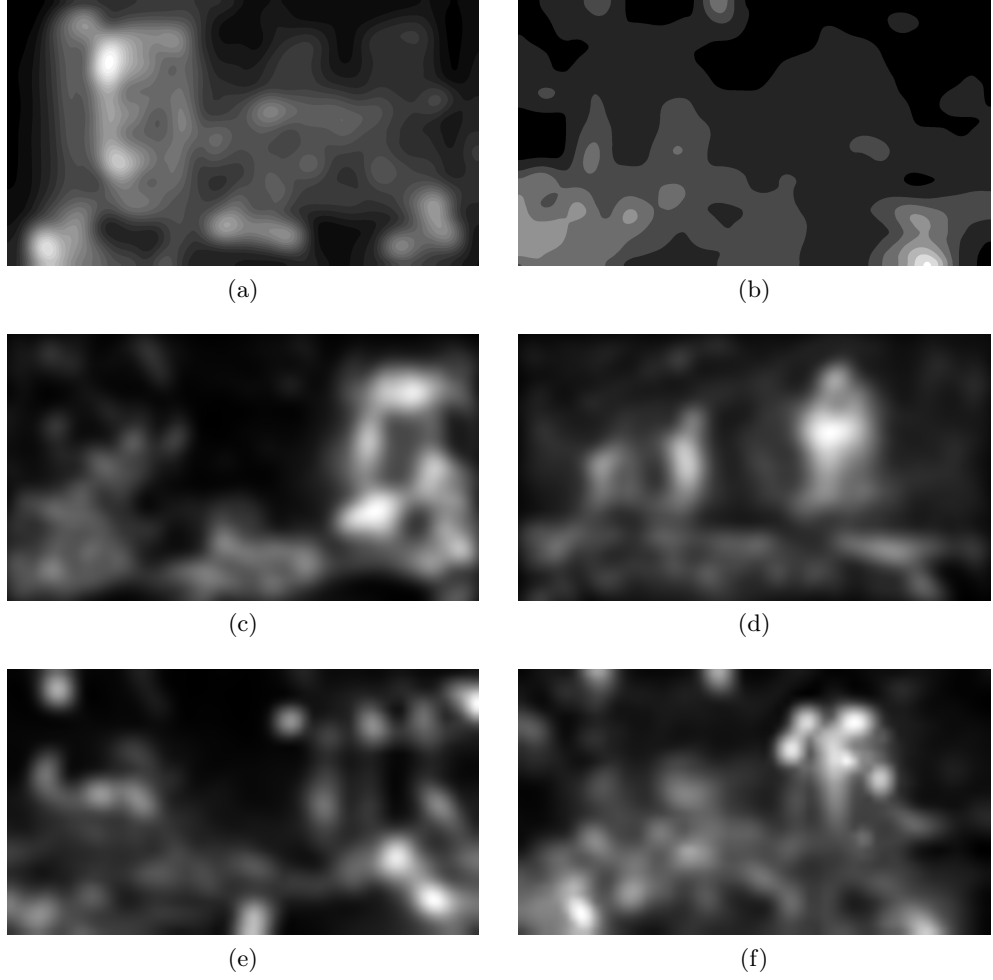


Figure 5.16: (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 22 image and right for No 24 image.

5.4 Saliency Guided HDR Visual Content Encoding

In this section, the proposed saliency-guided HDR image and video compressions are illustrated.

5.4.1 Image Encoding

Saliency-based compressions [49] utilize saliency map as the ROI. The proposed saliency-guided compression is able to segment the image into N layers by the saliency map. In the experiment, the input image is segmented into two-layers in order to avoid the overhead costs of extra layers. The number of encoding layers is not limited by two, it could be more than two depending

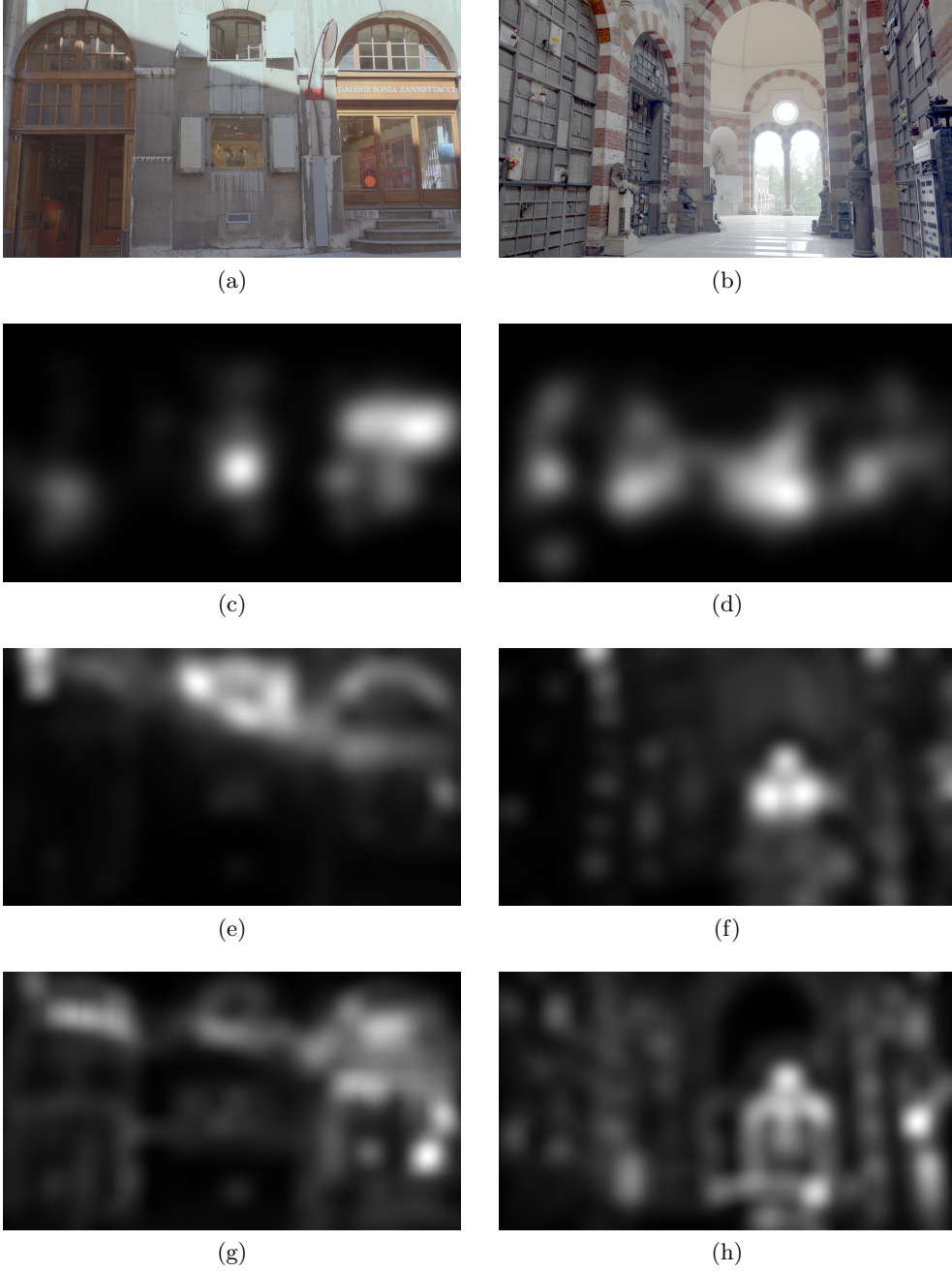


Figure 5.17: Tone mapped image (a, b), Saliency maps of (c, d) ground truth, (e, f) Itti's, (g, h) CF, from top to bottom. Column of left for No 33 image and right for No 39 image.

on the requirement. The threshold (t) of segmentation is set to a given value from 0 to 1. When prediction visual saliency is lower than the given threshold (t), a lower compression quality setting is applied; when saliency is larger than the given threshold (t), a higher compression quality setting is applied.

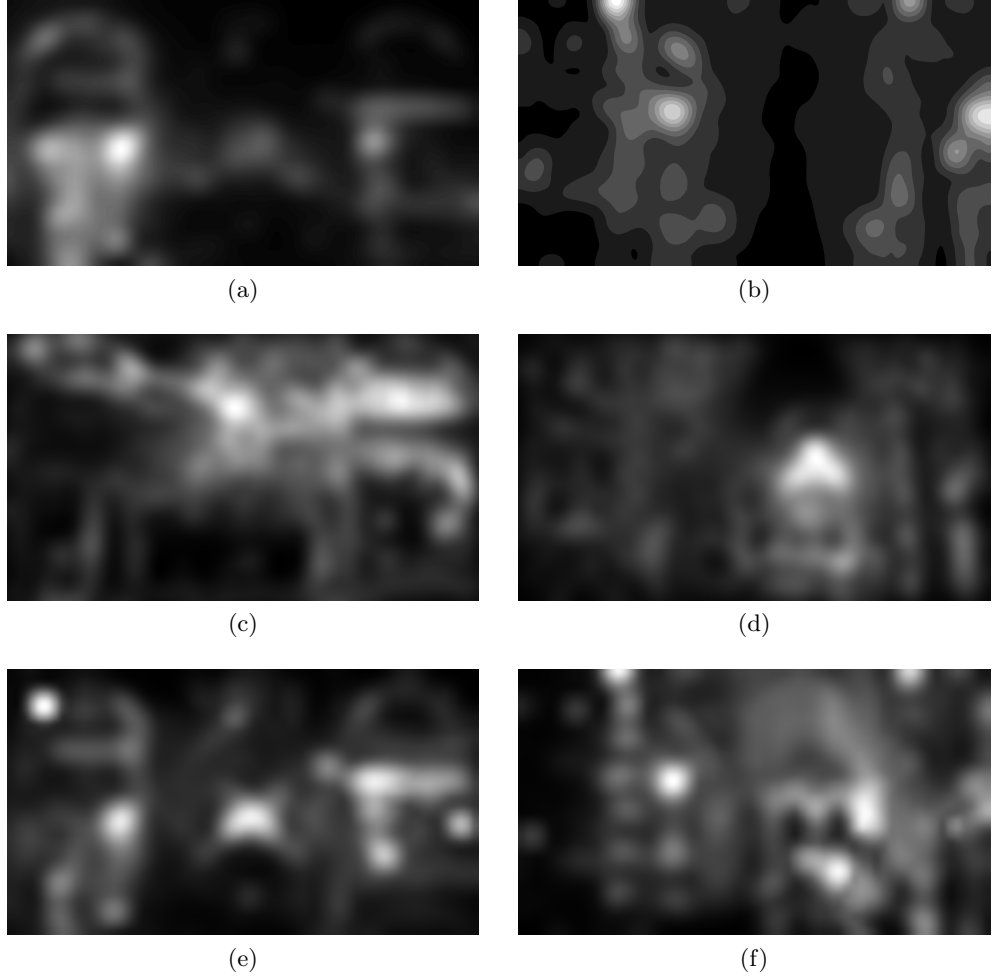


Figure 5.18: (a, b) Gao's, (c, d) Dong's and (e, f) proposed model from top to bottom. Column of left for No 33 image and right for No 39 image.

The example of proposed encoding is illustrated by Figure 5.19

$$M_t(x, y) = \begin{cases} 0, & S_M \leq t \\ 1, & S_M > t \end{cases}, \quad (5.11)$$

where the ROI mask M_t is the threshold t satisfy $\frac{\sum M_t}{R} = \frac{3}{10}$ of saliency map S_M and the R is the resolution of image.

Although the visual saliency threshold and compression quality are adjustable and independent, the salient regions which are larger than half of the image size will affect compression efficiency. The top 30% salient regions are

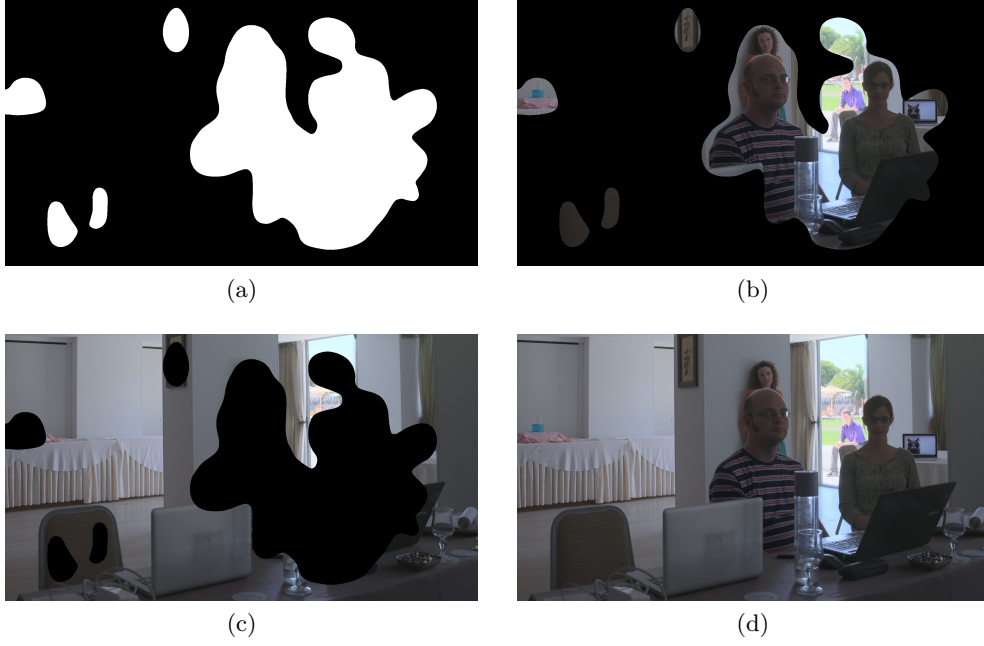


Figure 5.19: (a) The saliency mask for segmentation, (b) Salient regions after segmentation, (c) Non-Salient regions after segmentation. (d) The summation of decoded Salient regions and Non-Salient regions

used as the ROI mask.

Figure 5.20 shows the proposed model's decoding scheme. The proposed decoding process does not require the mask of the saliency map, because for each layer the QP is embedded in the bitstream, and there are only two-layers in total.

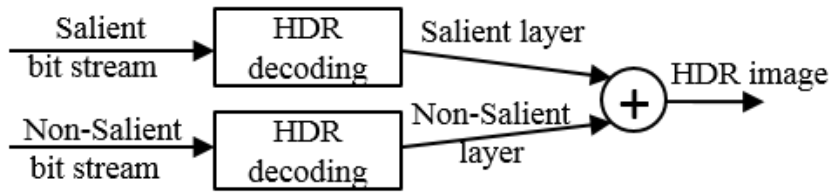


Figure 5.20: The decoding block diagram of the proposed model.

5.4.2 Video Encoding

The video encoding follows the same scheme of image encoding. It is worth mentioning that the proposed scheme aims to provide a salient and non-salient

Table 5.8: MSE and eye-weighted MSE

| Image | test image 1 | test image 2 |
|----------|------------------|------------------|
| | MSE | MSE |
| C1 607KB | 47 | 23 |
| C2 121KB | 1721 | 273 |
| C3 363KB | 606 | 254 |
| | eye-weighted MSE | eye-weighted MSE |
| C1 607KB | 124 | 38 |
| C2 121KB | 10867 | 629 |
| C3 363KB | 71 | 193 |

mask for encoding with compression quality depending on requirement. The quantization parameter (QP) of the salient region is fixed to equal to the normal compression counterpart to get a single RD curve of the proposed method (otherwise there are too many combinations), so the resulting quality of salient region will always be equal to the normal compression. The rate-distortion depends on the settings of compression quality of the salient region and non-salient region. Hence, only one video example is demonstrated to describe the proposed concept.

5.4.3 Experimental Results and Discussion

In this section, the experiment results are evaluated and discussed.

Li *et al.* [74] proposed an eye-weighted PSNR (EWPSNR) for quality measurement for saliency application. However, it requires the actual eye-fixation points, and these points are not available at most of the times, an example is given in Figure 5.21. The eye-weighted PSNR or MSE are calculated by the original MSE and PSNR maps multiplied by a map of Gaussian filtered eye-fixation points.

The results show that the saliency-guided compression has a higher eye-weighted MSE then MSE. However, the ground truth fixation points are not always available in all circumstances.

The traditional rate-distortion curve is not considered to be an evaluation method, because when only considering the salient regions, the quality of



Figure 5.21: The original image, fixation density map, fixation points, k-means clustered Gaussian weighted map. from left to right, and top to bottom.

saliency-guided compression is fixed by a certain QP factor. Followed by an example in Figure 5.22, the QPs for the salient region are fixed to 1, 5, and 10 separately while they are not fixed for non-salient regions. As shown in the rate-distortion results, the HDR-VDP of salient regions of saliency-guided compression is constant because the proposed method fixed the QP for salient regions. And the HDR-VDP final result is a global result rather than detection rate per each pixel. Therefore, the proposed method is demonstrated

by showing results of HDR-VDP map for the whole image.

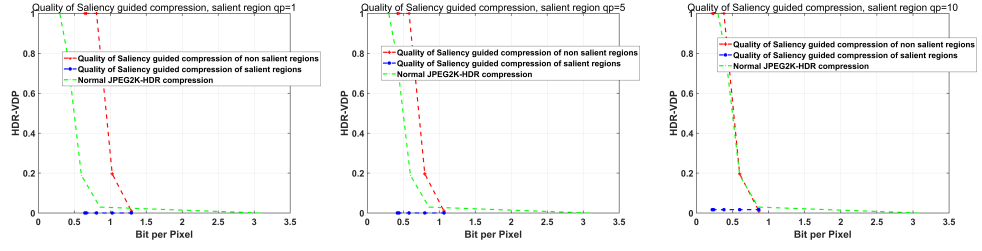


Figure 5.22: RD curves saliency guided compression.

5.4.3.1 The Costs of Saliency Map

As mentioned in Section 5.4.1, the proposed encoding could be more than two depends on the requirement, and it required more bits to store the QP of each layer. The cost of encoding a binary saliency map and a density saliency map can be approximate by the Shannon entropy:

$$H = - \sum_{i=1}^n p(x_i) \log p(x_i), \quad (5.12)$$

where p is the probability of each unique pixel value x_i , n is the total number of unique pixel values.

After evaluated around 100 of saliency maps, the average encoding cost of a binary saliency map and a density saliency map are approximately 1 bit per pixel, and 3 bit per pixel respectively. Since the proposed method is two-layers coding, the above extra encoding cost can be saved, however, the encoding and decoding procedures are repeated to each layers.

5.4.3.2 Image Results

The compression ratio of 20 and 100 are chosen because of the relative differences is more distinguish in the colour map of HDR-VDP.

Figure 5.23 shows two examples of the HDR image compression application guided by visual saliency map. The distortions of JPEG2000-HDR compression are measured by HDR-VDP metric in rows (1) – (3) in Figure 5.24. The

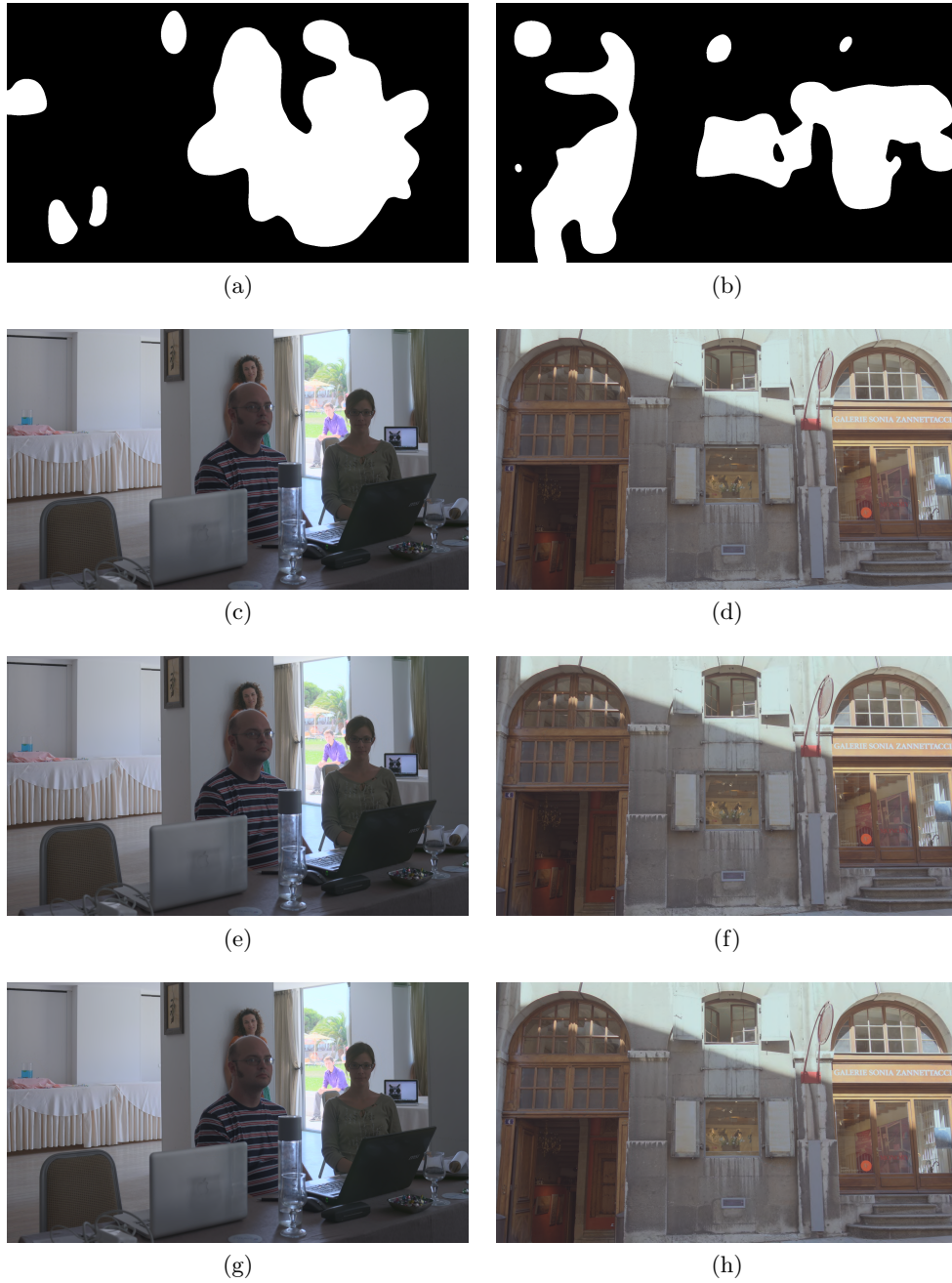


Figure 5.23: Results of proposed HDR compression application, (a, b) Segmentation by threshold, (c, d) Decoded image of JPEG2K-HDR – the compressed file size are 607 KB, when compression ratio is 20, (e, f) Decoded image of Visual Saliency guided JPEG2K-HDR – the compressed file size are 363 KB, when compression ratio is 20 for salient region, and 100 for non-salient region, (g, h) Decoded image of JPEG2K-HDR – the compressed file size are 121 KB, when compression ratio is 100

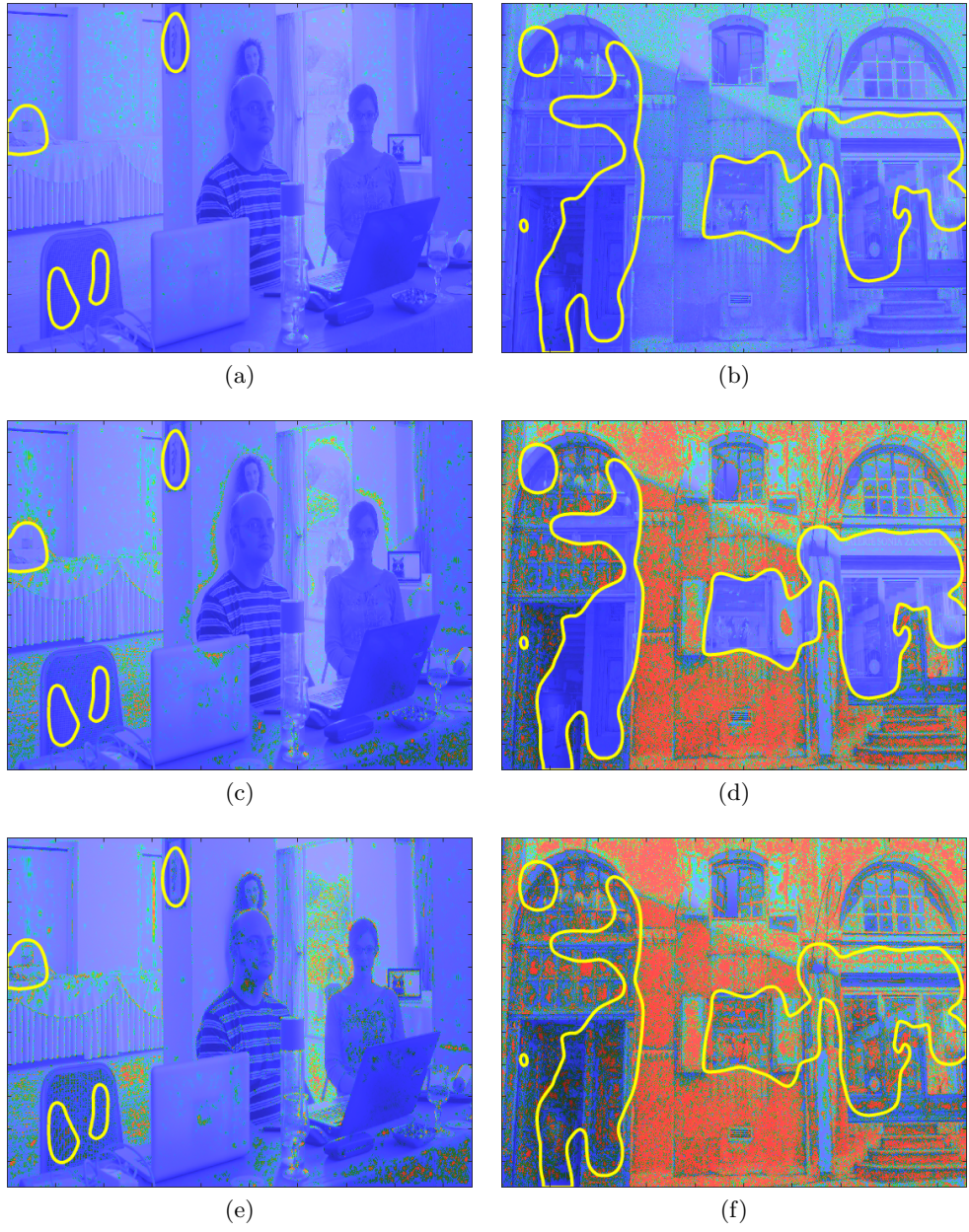


Figure 5.24: Results of proposed HDR compression application, (a, b) HDR-VDP map of decoded image of JPEG2K-HDR, (c, d) HDR-VDP map of decoded image of Visual Saliency guided JPEG2K-HDR, (e, f) HDR-VDP map of decoded image of JPEG2K-HDR for compression ratio 20.

saliency-guided JPEG2000-HDR compression obtained higher compression ratios while it keeps the same quality for salient regions. The colourmap used in HDR-VDP prediction maps is showed in Figure 5.27, red colour means the highest error and blue colour means the lowest error.



Figure 5.25: Results of proposed HDR compression application, (a, b) Segmentation by threshold, (c, d) Decoded image of JPEG2K-HDR – the compressed file size are 607 KB, when compression ratio is 20, (e, f) Decoded image of Visual Saliency guided JPEG2K-HDR – the compressed file size are 363 KB, when compression ratio is 20 for salient region, and 100 for non-salient region, (g, h) Decoded image of JPEG2K-HDR – the compressed file size are 121 KB, when compression ratio is 100

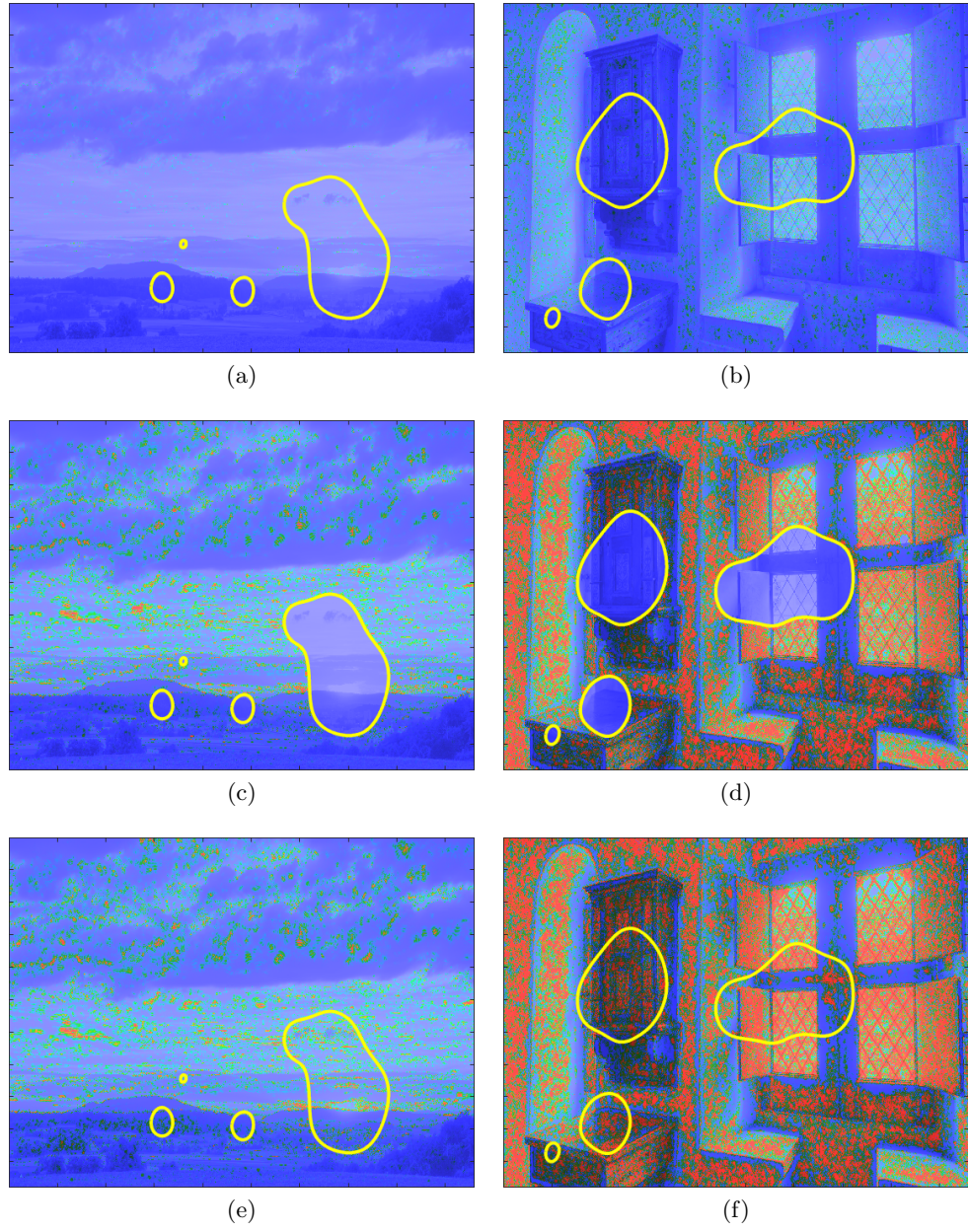


Figure 5.26: Results of proposed HDR compression application, (a, b) HDR-VDP map of decoded image of JPEG2K-HDR, (c, d) HDR-VDP map of decoded image of Visual Saliency guided JPEG2K-HDR, (e, f) HDR-VDP map of decoded image of JPEG2K-HDR for compression ratio 20.

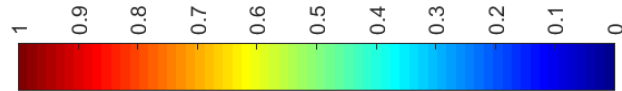


Figure 5.27: The colourmap used for visualizing HDR-VDP metric.

To have a better demonstration of the proposed method, the ROI mask is generated by the proposed model. The accurate saliency map leads to the efficient saliency-guided compression.

5.4.3.3 Video Results

As shown in Figure 5.28, the quality of the salient regions of saliency-guided Mapping-free Motion JPEG2000 compression is always equal to the high-quality normal compression. If the quality is evaluated by the non-salient regions, the quality of saliency-guided compression is always equal to the low-quality normal compression. The final rate-distortion depends on the differences between high and low quality compressions, and it also depends on the size of the ROI mask. Figure 5.28 shows that the qualities (HDR-VDP, MSE) of the salient region of the proposed method are always the same compared to the normal high-quality compression.

Figure 5.29 shows the Mapping-free HEVC two fixed QP 0 and 10 settings for salient regions, while the settings are 0,10,20,30 and 40 for the non-salient regions. However, the experimental results suggest that the HDR-VQM could not discriminate very well of the distortion between QP 0 and 10, the normalized results are 0.869 and 0.871 respectively (the HDR-VQM suggested that the quality of QP 10 encoded video is better than QP 0 which is wrong), and the MSE and HDR-VDP could discriminate these distortions, as shown in Figure 5.30. The HDR-VQM per frame result without temporal polling is displayed in the Figure 5.31.

The proposed saliency guided N layers compression is evaluated in two-layers, and the costs (around 1 bit per pixel according to the experimental results) of sending a saliency map (QP map) to the decoder is avoided, since the QP is embedded in each bitstream. The shortcoming of the proposed method is the additional computation costs for the extra layer for both encoding and decoding.

The proposed saliency guided HDR compression scheme can be applied to

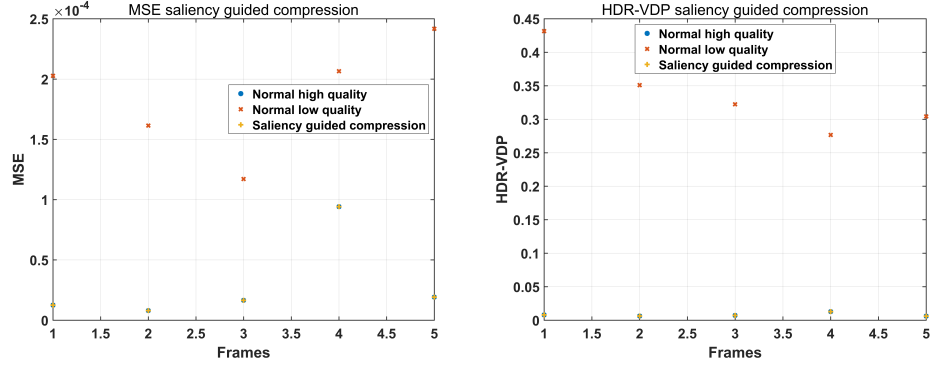


Figure 5.28: MSE and HDR-VDP of saliency guided JPEG2000-Motion compression.

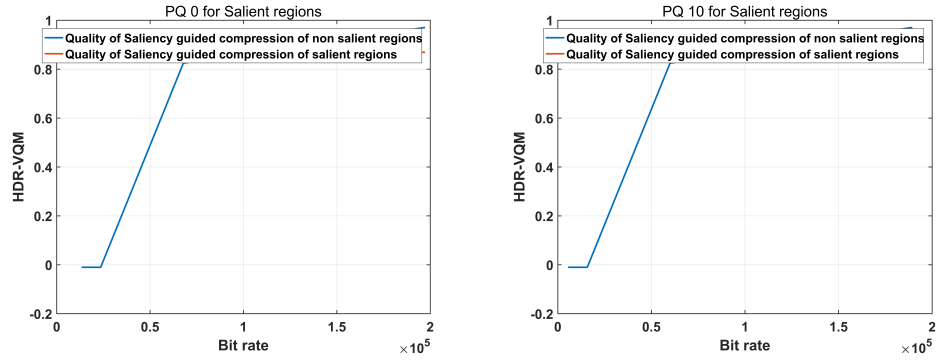


Figure 5.29: HDR-VQM of saliency guided HEVC HDR Mapping-free compression.

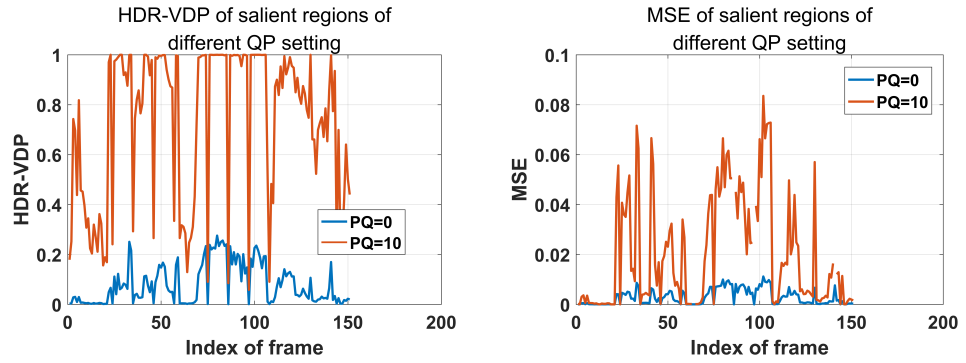


Figure 5.30: HDR-VDP and MSE of saliency guided HEVC HDR Mapping-free compression.

every non-saliency guided HDR compressions including the proposed method in Chapter 3 and Chapter 4. The average bit-rate savings are at least 30% if the QP setting of the non-salient layer is at least double of the salient layer,

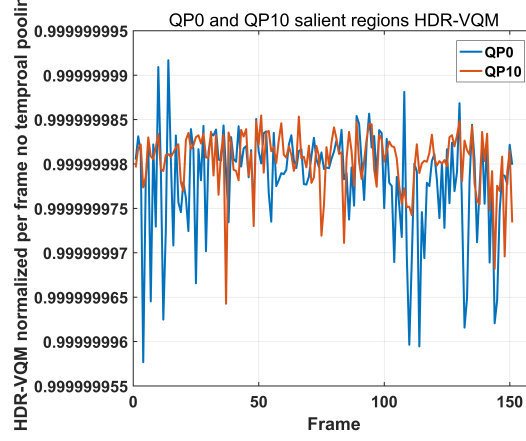


Figure 5.31: HDR-VQM per frame results of the orange lanes in Figure 5.29 which are the salient layers of QP0 and QP10.

and only use top 30% salient regions as the ROI mask. However, the current objective visual quality metrics fail to weigh the distortions by visual saliency. The HVS driven metrics tend to weigh mid-frequency error differently than the low and high-frequency errors, but it is not the particular visual saliency approach. The eye-weighted PNSR can be extended to these objective metrics, however, it requires the actual eye-fixation points from an eye-tracker. Hence, the HDR-VDP pixel-wise probability map is used to describe the proposed visual saliency-guided HDR compression. Furthermore, experimental results suggest that the HDR-VQM metric could not discriminate the distortions well between QP 0 and 10 of HEVC encoding.

The proposed scheme does not include bit-allocation or optimal QP estimation algorithm, because of the extra complexity required and the visual perception of the HDR contents depend on the view conditions. With defined viewing condition and additional computational power and desired bit-rate, an optimization can be applied for the two-layers methods to optimize the bit-allocation between the salient layer and non-salient layer for a given bit-rate.

5.5 Summary

In this chapter, the shortcomings of the current HDR visual saliency models, and HDR visual saliency-guided compression are explored. The visual saliency model for the HDR image is proposed, and the performance is evaluated against four existing methods. The results show that the proposed model has the highest mean of PCC and KLD to the ground truth eye tracking data. For compression application, the saliency prediction map could be used to determine the ROI. An accurate visual saliency prediction could lead to efficient HDR image compression and evaluation. Furthermore, the performances of temporal saliency models are evaluated. The results prove that the additional temporal saliency predicted alone by the temporal model is negligible, and the 93% of temporal saliency is already predicted by the proposed spatial model. The proposed HDR image and video compression schemes are validated by HDR-VDP metric versus bit-rate. The proposed compression scheme uses the high-quality compression setting to compress the salient regions and low-quality compression setting to compress the non-salient regions. Compared to the normal compressions, the qualities of the proposed method are always the same for the salient region when the quality settings are the same. The coding gain is obtained by the low-quality compression setting for the non-salient regions.

Chapter 6

Conclusions

Recent years have seen the rapid development of HDR content. The HDR encodings are called scene-referred standard, which intends to represent the actual physical value of the light rather than exposed value. However, the storage and transmission cost of the HDR contents is higher than LDR 8 bits representation. The floating-points of HDR value require at most 64 bits per colour channel, which is 8 times of conventional LDR imaging. Hence the compression of HDR is mandatory. The aim of this thesis is to address challenges in compressing HD HDR images so that they can be efficiently stored and transmitted.

In Chapter 2, three research questions were identified by reviewing and evaluating the current HDR format, the TMOs for LDR backward compatibility, the state-of-the-art HDR compression methods, the HDR visual quality metrics and the visual saliency models. They are: 'What is the relationship of DWT lifting steps and TMO'; 'What is the importance of the TMO in the HDR compression'; 'What are the shortcomings of the current HDR visual saliency models, and How to compress HDR content by visual-saliency'.

In Chapter 3, the relationship of DWT lifting steps and TMO are explored, by proposing a new compression scheme for HDR images based on DWT lifting with TMO built-in. The demonstration starts from 1D Haar lifting, and extends to 2D Haar, 2D 3/5 and 2D 5/3 liftings. The properties of global TMO

and ITMO integrating to the DWT liftings are introduced. This approach achieved an equivalent performance of MSE and HDR-VDP compared to existing HDR image compression methods, with LDR backward compatibility. At the same bit-rate, the quality differences among Haar, 3/5, 5/3 and 9/7 liftings are noticeable, because of the characteristics of the different wavelets *i.e.* vanishing moment. Therefore, a fine selection of wavelet can lead to a better compression performance. The compression quality can be determined by the HDR-VDP metric with given view condition. The proposed scheme can be extended to HDR video compression, the flickering artifacts caused by TMO can be reduced by a temporal filtering of TMO brightness parameters. Based on the evaluation results of this chapter and with the development of HDR displays, the experimental results suggest that the HDR compression schemes with LDR backward compatibility increase the overall complexity of system, also, the TMOs cause artifacts in the LDR sequences. Moreover, the LDR backward compatibility can always be obtained by a TMO after fully decoding the bit-stream. The proposed method is not robust for the extreme dynamic range cases, because of the cross-LDR-HDR subbands decoding.

In Chapter 4, the importance of the TMO in the HDR compression is evaluated, by proposing a mapping-free post-HDR image compression based on JPEG and JPEG2000 standard codecs for RGBE, EXR and PFM HDR image formats. The quantization errors affected by TMO and the inverse tone-mapped errors of the quantization errors when decoding the HDR content are eliminated in the proposed method. The results of mapping-free post HDR image compression of RGBE and OpenEXR images are compared with two post HDR image compression methods: JPEG-HDR and JPEG2000-HDR. The proposed mapping-free post-HDR image compression scheme has the lowest computational complexity (50% lower computational complexity than the previous) among all methods, because the tone-mapping or logarithmic transform step is skipped in the proposed method. The compression distortion of the proposed method is comparable to JPEG-HDR and JPEG2000-HDR in

MSE and HDR-VDP at low compression ratios. The proposed mapping-free compression scheme can be extended to video compression. The experimental results illustrate that the proposed method obtained equivalent compression performance with low computational complexity. The proposed method does not exploit the low-level features of legacy codecs, the compression performances of legacy codecs are not improved.

In Chapter 5, the shortcomings of the current HDR visual saliency models are overcome and HDR visual saliency-guided compression is explored, by proposing visual saliency model for the HDR image. The proposed model is compared with four existing methods and has the highest mean of PCC and KLD to the ground truth eye tracking data, it outperforms others by 10% for spatial visual prediction task with 70% lower computational complexity. The saliency prediction map provides guidance of the ROI. An accurate visual attention prediction offers efficient HDR image compression and evaluation. The evaluation prove that the additional temporal saliency predicted alone by the temporal model is negligible, and the 93% of temporal saliency is already predicted by the proposed spatial model. The proposed HDR image and video compression schemes are validated by HDR-VDP metric versus bitrate. The experimental results also suggest that the HDR-VQM metric could not discriminate the distortions between PQ 0 and 10 of HEVC encoding. The proposed compression scheme uses high-quality compression setting to compress the salient regions and low-quality compression setting to compress the non-salient regions. Compared to the normal compressions, the qualities of the proposed method are always the same for the salient region when the quality settings are the same. The coding gain is obtained by the low-quality compression setting for the non-salient regions. The specific bit-allocation can be suggested by the subjective evaluation.

In conclusion, employing high-quality compression setting for salient regions and low-quality compression setting for non-salient regions leads to coding gains. The proposed saliency-guided HDR visual content encoding can be

employed to current HDR encodings, the proposed DWT-TMO and Mapping-Free HDR encodings. Hence, a lower compression bit-rate can be achieved while maintaining the same perceptual visual quality.

6.1 Future work

This thesis presents an HVS saliency-guided low complexity HDR visual content encoding scheme. The shortcomings and limitations can be further investigated.

For the proposed schemes, the subjective validation experiments can be appreciated. Since there is only one HDR video saliency ground truth available, the proposed scheme can be further validated by a complete HDR video saliency dataset. Some recent glasses type of eye-tracking products can record the LDR as the reference map (event log) through a camera, while the fixation and gaze path are coming from the real world scene. A machine learning algorithm could map between LDR recorded sequence and HDR fixations. A new model of HDR saliency can be realized.

A further research could be done for improving existing HDR objective evaluation metrics. The spatial model of HDR-VDP is more accurate than the HDR-VQM, and there is no temporal pooling rule for HDR-VDP. For both of metrics, the effect of chrominance distortions is not evaluated directly.

The proposed algorithms can be implemented in more efficiently way by using multi-core CPUs and graphics processing units (GPUs) to accelerate computation. The parallel computing make it possible to process images and videos in real-time.

For the moment, the acquisition of the HDR content is still the main limitation. Even through there are solutions, they are expensive in both price and computational complexity. A low cost and computational complexity HDR acquisition solution, especially for video, can provide more content for the related research.

References

- [1] “IEEE Standard for Floating-Point Arithmetic,” *IEEE Std 754-2008*, pp. 1–70, Aug 2008.
- [2] C. Abhayaratne, “Reversible wavelet filter banks with side information-less spatially adaptive low-pass filters,” *Journal of Electronic Imaging*, vol. 20, no. 3, pp. 033 012–033 012, 2011.
- [3] A. Artusi, R. Mantiuk, T. Richter, P. Korshunov, P. Hanhart, T. Ebrahimi, and M. Agostinelli, “JPEG XT: A Compression Standard for HDR and WCG Images [Standards in a Nutshell],” *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 118–124, 2016.
- [4] T. O. Aydin, R. Mantiuk, and H.-P. Seidel, “Extending quality metrics to full luminance range images,” in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 68 060B–68 060B.
- [5] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, “Dynamic range independent image quality assessment,” in *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3. ACM, 2008, p. 69.
- [6] T. O. Aydin, N. Stefanoski, S. Croci, M. Gross, and A. Smolic, “Temporally coherent local tone mapping of HDR video,” *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, p. 196, 2014.
- [7] A. Banitalebi-Dehkordi, M. Azimi, M. T. Pourazad, and P. Nasiopoulos, “Compression of High Dynamic Range video using the HEVC and

- H.264/AVC standards,” in *10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness*, Aug 2014, pp. 8–12.
- [8] A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos, “A learning-based visual saliency fusion model for High Dynamic Range video,” in *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE, 2015, pp. 1541–1545.
- [9] D. Bhowmik, M. Oakes, and C. Abhayaratne, “Visual attention-based image watermarking,” *IEEE Access*, vol. 4, pp. 8002–8018, 2016.
- [10] L. Bilodeau and J. Faubert, “Isoluminance and chromatic motion perception throughout the visual field,” *Vision research*, vol. 37, no. 15, pp. 2073–2081, 1997.
- [11] G. Bjontegaard, “Calculation of average PSNR differences between RD-Curves,” *ITU-T SG16 Q*, vol. 6, p. 35, 2001.
- [12] —, “Improvements of the bd-psnr model,” *ITU-T SG16 Q*, vol. 6, p. 35, 2008.
- [13] R. Boitard, “HDR temporal brightness artifacts database,” <http://www.ece.ubc.ca/~rboitard/>, Sept 2016.
- [14] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch, “Survey of temporal brightness artifacts in video tone mapping,” in *HDRi2014-Second International Conference and SME Workshop on HDR imaging*. Cite-seer, 2014.
- [15] A. Borji and L. Itti, “State-of-the-art in visual attention modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 35, no. 1, pp. 185–207, 2013.
- [16] A. P. Bradley and F. W. Stentiford, “JPEG 2000 and region of inter-

est coding,” in *Digital Image Computing Techniques and Applications*, vol. 2, 2002, pp. 1–6.

- [17] R. Brémond, J. Petit, and J.-P. Tarel, “Saliency maps of High Dynamic Range images,” in *Trends and Topics in Computer Vision*. Springer, 2010, pp. 118–130.
- [18] F. Campbell, J. Kulikowski, and J. Levinson, “The effect of orientation on the visual resolution of gratings,” *The Journal of Physiology*, vol. 187, no. 2, p. 427, 1966.
- [19] M. Chen, G. Qiu, Z. Chen, and C. Wang, “JPEG compatible coding of high dynamic range imagery using tone mapping operators,” in *Proc. 25th Picture Coding Symposium*, 2006.
- [20] C. Christopoulos, A. Skodras, and T. Ebrahimi, “The JPEG2000 still image coding system: an overview,” *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103–1127, 2000.
- [21] S. J. Daly, “Visible differences predictor: an algorithm for the assessment of image fidelity,” in *SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1992, pp. 2–15.
- [22] I. Daubechies, “Orthonormal bases of compactly supported wavelets,” *Communications on Pure and applied Mathematics*, vol. 41, no. 7, pp. 909–996, 1988.
- [23] I. Daubechies and W. Sweldens, “Factoring wavelet transforms into lifting steps,” *Journal of Fourier Analysis and Applications*, vol. 4, no. 3, pp. 247–269, 1998.
- [24] P. E. Debevec and J. Malik, “Recovering High Dynamic Range radiance maps from photographs,” in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, ser.

SIGGRAPH '97. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1997, pp. 369–378. [Online]. Available: <http://dx.doi.org/10.1145/258734.258884>

- [25] P. Debevec, “Recovering High Dynamic Range radiance maps from photographs,” in *Proc. SIGGRAPH'97*, 1997.
- [26] Y. Dong, M. T. Pourazad, and P. Nasiopoulos, “Human Visual System-based saliency detection for High Dynamic Range content,” *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 549–562, April 2016.
- [27] Y. Dong, E. Nasiopoulos, M. T. Pourazad, and P. Nasiopoulos, “High dynamic range video eye tracking dataset,” in *Proc. 2nd Int. Conf. Electron., Signal Process. Commun.*, 2014, pp. 56–59.
- [28] —, “High Dynamic Range video eye tracking dataset,” in *2nd International Conference on Electronics, Signal processing and Communications, Athens*, 2015.
- [29] Y. Dong, P. Nasiopoulos, and M. T. Pourazad, “Hdr video compression using high efficiency video coding (hevc),” *reproduction*, vol. 6, 2012.
- [30] Y. Dong, M. T. Pourazad, and P. Nasiopoulos, “Human Visual System-based saliency detection for High Dynamic Range content,” *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 549–562, 2016.
- [31] F. Drago, W. L. Martens, K. Myszkowski, and H.-P. Seidel, *Perceptual evaluation of tone mapping operators with regard to similarity and preference*. Citeseer, 2002.
- [32] F. Drago, K. Myszkowski, T. Annen, and N. Chiba, “Adaptive logarithmic mapping for displaying high contrast scenes,” in *Computer Graphics Forum*, vol. 22, no. 3. Wiley Online Library, 2003, pp. 419–426.
- [33] F. Dufaux, G. J. Sullivan, and T. Ebrahimi, “The JPEG XR image cod-

ing standard [standards in a nutshell],” *IEEE Signal Processing Magazine*, vol. 26, no. 6, pp. 195–199, 204–204, November 2009.

- [34] F. Dufaux, P. Le Callet, R. Mantiuk, and M. Mrak, *High Dynamic Range Video: From Acquisition, to Display and Applications*. Academic Press, 2016.
- [35] F. Durand and J. Dorsey, “Fast bilateral filtering for the display of high-dynamic-range images,” in *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3. ACM, 2002, pp. 257–266.
- [36] G. Eilertsen, R. K. Mantiuk, and J. Unger, “Real-time noise-aware tone mapping,” *ACM Transactions on Graphics (TOG)*, vol. 34, no. 6, p. 198, 2015.
- [37] R. Fattal, D. Lischinski, and M. Werman, “Gradient domain high dynamic range compression,” in *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3. ACM, 2002, pp. 249–256.
- [38] T. Feng, S. Pal, and C. Abhayaratne, “Mapping-free high dynamic range image compression,” in *Proceedings of the 2015 IEEE 5th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*. IEEE, 2015, pp. 172–175.
- [39] R. P. Feynman, R. B. Leighton, and M. L. Sands, *The Feynman lectures on physics: Mainly mechanics, radiation, and heat*. Basic Books, 2011, vol. 1.
- [40] D. Flynn, D. Marpe, M. Naccari, T. Nguyen, C. Rosewarne, K. Sharman, J. Sole, and J. Xu, “Overview of the Range Extensions for the HEVC standard: Tools, profiles, and performance,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 4–19, 2016.

- [41] R. Freedman, R. Geller, and W. J. Kaufmann, *Universe: The solar system*. Macmillan, 2010.
- [42] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, “Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays,” in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90 230X–90 230X.
- [43] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, “Saliency-based video segmentation with graph cuts and sequentially updated priors,” in *2009 IEEE International Conference on Multimedia and Expo*. IEEE, 2009, pp. 638–641.
- [44] X. Gao, S. Brooks, and D. V. Arnold, “Automatic blended tone mapping through evolutionary optimization,” in *2016 IEEE Congress on Evolutionary Computation (CEC)*, July 2016, pp. 3855–3862.
- [45] —, “Virtual photograph based saliency analysis of High Dynamic Range images,” in *Proceedings of the Symposium on Computational Aesthetics*. ACM, 2013, pp. 87–92.
- [46] —, “Automated parameter tuning for tone mapping using visual saliency,” *Computers & Graphics*, vol. 52, pp. 171–180, 2015.
- [47] R. C. Gonzalez and R. E. Woods, “Digital image processing,” *Nueva Jersey*, 2008.
- [48] M. Grundmann, V. Kwatra, M. Han, and I. Essa, “Efficient hierarchical graph-based video segmentation,” in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 2141–2148.
- [49] C. Guo and L. Zhang, “A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression,”

IEEE Transactions on Image Processing, vol. 19, no. 1, pp. 185–198, Jan 2010.

- [50] H. Hadizadeh and I. V. Bajić, “Saliency-aware video compression,” *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 19–33, 2014.
- [51] P. Hanhart, M. V. Bernardo, P. Korshunov, M. Pereira, A. M. Pinheiro, and T. Ebrahimi, “HDR image compression: a new challenge for objective quality metrics,” in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2014, pp. 159–164.
- [52] —, “HDR image compression: A new challenge for objective quality metrics,” in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept 2014, pp. 159–164.
- [53] P. Hanhart, L. Krasula, P. Le Callet, and T. Ebrahimi, “How to Benchmark Objective Quality Metrics from Paired Comparison Data?” in *8th International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.
- [54] P. Hanhart, M. Rerabek, and T. Ebrahimi, “Subjective and Objective Evaluation of HDR Video Coding Technologies,” 2016.
- [55] J. Harel, C. Koch, P. Perona *et al.*, “Graph-based visual saliency,” in *NIPS*, vol. 1, no. 2, 2006, p. 5.
- [56] L. Itti, “Automatic foveation for video compression using a neurobiological model of visual attention,” *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004.
- [57] L. Itti, N. Dhavale, and F. Pighin, “Realistic avatar eye and head animation using a neurobiological model of visual attention,” in *Optical science and technology, SPIE’s 48th annual meeting*. International Society for Optics and Photonics, 2004, pp. 64–78.

- [58] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, no. 11, pp. 1254–1259, 1998.
- [59] M. Iwahashi and H. Kiya, “Two layer lossless coding of HDR images,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2013, pp. 1340–1344.
- [60] T. Jinno, M. Okuda, and N. Adami, “New local tone mapping and two-layer coding for HDR images,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2012, pp. 765–768.
- [61] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in *2009 IEEE 12th international conference on Computer Vision*. IEEE, 2009, pp. 2106–2113.
- [62] K. Kabaja, “Storing of High Dynamic Range images in JPEG/JFIF files,” *Proc. of the CESC G*, pp. 1–8, 2005.
- [63] F. Kainz and R. Bogart, “Technical introduction to OpenEXR,” *Industrial Light and Magic*, 2009.
- [64] H. Kikuchi, T. Deguchi, and M. Okuda, “Lossless compression of logluv32 hdr images by simple bitplane coding,” in *Picture Coding Symposium (PCS), 2013*. IEEE, 2013, pp. 265–268.
- [65] C. Kiser and E. Reinhard, “Real time automated tone mapping system for HDR video,” in *IEEE International Conference on Image Processing*. IEEE, 2012, pp. 2749–2752.
- [66] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, “Subjective quality assessment database of hdr images compressed with JPEG XT,” in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2015, pp. 1–6.

- [67] G. W. Larson, “Logluv encoding for full-gamut, high-dynamic range images,” *Journal of Graphics Tools*, vol. 3, no. 1, pp. 15–31, 1998.
- [68] G. W. Larson, H. Rushmeier, and C. Piatko, “A visibility matching tone reproduction operator for high dynamic range scenes,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.
- [69] S. Lasserre, F. Le Léanec, T. Poirier, and F. Galpin, “Backward compatible hdr video compression system,” in *Data Compression Conference (DCC), 2016*. IEEE, 2016, pp. 309–318.
- [70] O. Le Meur and T. Baccino, “Methods for comparing scanpaths and saliency maps: strengths and weaknesses,” *Behavior Research Methods*, vol. 45, no. 1, pp. 251–266, 2013.
- [71] S.-H. Lee, J.-W. Kang, and C.-S. Kim, “Compressed domain video saliency detection using global and local spatiotemporal features,” *Journal of Visual Communication and Image Representation*, vol. 35, pp. 169–183, 2016.
- [72] A. G. Leventhal, *The neural basis of visual function*. CRC Pr I Llc, 1991, vol. 4.
- [73] Y. Li, L. Sharan, and E. H. Adelson, “Compressing and companding High Dynamic Range images with subband architectures,” in *ACM transactions on graphics (TOG)*, vol. 24, no. 3. ACM, 2005, pp. 836–844.
- [74] Z. Li, S. Qin, and L. Itti, “Visual attention guided bit allocation in video compression,” *Image and Vision Computing*, vol. 29, no. 1, pp. 1–14, 2011.
- [75] I. Light and Magic, “High Dynamic Range image examples in openexr formats,” <http://www.openexr.com/downloads.html>, Sept 2016.

- [76] W.-C. Lin and Z.-C. Yan, “Attention-based High Dynamic Range imaging,” *The Visual Computer*, vol. 27, no. 6-8, pp. 717–727, 2011.
- [77] L. Liu and G. Fan, “A new JPEG2000 region-of-interest image coding method: Partial significant bitplanes shift,” *IEEE Signal Processing Letters*, vol. 10, no. 2, pp. 35–38, 2003.
- [78] A. Lüscho and H. Nothdurft, “Pop-out of orientation but no pop-out of motion at isoluminance,” *Vision Research*, vol. 33, no. 1, pp. 91–104, 1993.
- [79] D. Mahapatra, S. O. Gilani, and M. K. Saini, “Coherency based spatio-temporal saliency detection for video object segmentation,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 3, pp. 454–462, 2014.
- [80] D. Mahapatra, S. Winkler, and S.-C. Yen, “Motion saliency outweighs other low-level features while watching videos,” in *Electronic Imaging 2008*. International Society for Optics and Photonics, 2008, pp. 68 060P–68 060P.
- [81] Z. Mai, H. Mansour, R. Mantiuk, P. Nasiopoulos, R. Ward, and W. Heidrich, “Optimizing a tone curve for backward-compatible High Dynamic Range image and video compression,” *IEEE transactions on image processing*, vol. 20, no. 6, pp. 1558–1571, 2011.
- [82] Z. Mai, P. Nasiopoulos, and R. K. Ward, “Computationally efficient tone-mapping of high-bit-depth video in the ycbcr domain,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 1505–1508.
- [83] S. Mann and R. Picard, “Being ‘undigital’ with digital cameras: Extending Dynamic Range by combining differently exposed pictures,” M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts, Tech.

Rep. 323, 1994, also appears, IS&T's 48th annual conference, Cambridge, Massachusetts, May 1995.

- [84] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 68:1–68:10, Aug. 2008. [Online]. Available: <http://doi.acm.org/10.1145/1360612.1360667>
- [85] R. Mantiuk, S. J. Daly, K. Myszkowski, and H.-P. Seidel, "Predicting visible differences in high dynamic range images: model and its calibration," in *Electronic Imaging 2005*. International Society for Optics and Photonics, 2005, pp. 204–214.
- [86] R. Mantiuk, A. Efremov, K. Myszkowski, and H.-P. Seidel, "Backward compatible high dynamic range MPEG video compression," in *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 713–723.
- [87] R. Mantiuk, G. Krawczyk, K. Myszkowski, and H.-P. Seidel, "Perception-motivated high dynamic range video encoding," in *ACM Transactions on Graphics (TOG)*, vol. 23, no. 3. ACM, 2004, pp. 733–741.
- [88] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 40.
- [89] M. W. Marcellin, *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards, and Practice*. springer, 2002, vol. 1.
- [90] D. Marpe, T. Wiegand, and S. Gordon, "H. 264/MPEG4-AVC fidelity range extensions: Tools, profiles, performance, and application areas," in

- ICIP 2005. IEEE International Conference on Image Processing*, vol. 1. IEEE, 2005, pp. I–593.
- [91] D. Marpe, T. Wiegand, and G. J. Sullivan, “The h. 264/MPEG4 advanced video coding standard and its applications,” *IEEE Communications Magazine*, vol. 44, no. 8, pp. 134–143, 2006.
- [92] G. Mather, *Foundations of perception*. Taylor & Francis, 2006.
- [93] T. Mauthner, H. Possegger, G. Waltner, and H. Bischof, “Encoding based saliency detection for videos and images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2494–2502.
- [94] A. Motra and H. Thoma, “An adaptive logluv transform for high dynamic range video compression,” in *2010 17th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2010, pp. 2061–2064.
- [95] J. Munkberg, P. Clarberg, J. Hasselgren, and T. Akenine-Möller, “High dynamic range texture compression for graphics hardware,” in *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3. ACM, 2006, pp. 698–706.
- [96] M. Narwaria, M. P. Da Silva, and P. Le Callet, “HDR-VQM: An objective quality measure for high dynamic range video,” *Signal Processing: Image Communication*, vol. 35, pp. 46–60, 2015.
- [97] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi, “Visual attention in LDR and HDR images,” in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015.
- [98] M. Oakes, “Attention driven solutions for robust digital watermarking within media,” Ph.D. dissertation, The University of Sheffield, 2014.

- [99] M. Oakes and C. Abhayaratne, “Visual saliency estimation for video,” in *2012 13th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*. IEEE, 2012, pp. 1–4.
- [100] M. Okuda and N. Adami, “Effective color space representation for wavelet based compression of hdr images,” in *2007. ICIAP 2007. 14th International Conference on Image Analysis and Processing*. IEEE, 2007, pp. 388–392.
- [101] N. Ouerhani, J. Bracamonte, H. Hugli, M. Ansorge, and F. Pellandini, “Adaptive color image compression based on visual attention,” in *2001. Proceedings. 11th International Conference on Image Analysis and Processing*. IEEE, 2001, pp. 416–421.
- [102] E. Peli, “Contrast in complex images,” *JOSA A*, vol. 7, no. 10, pp. 2032–2040, 1990.
- [103] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still image data compression standard*. Springer Science & Business Media, 1992.
- [104] D. N. Recommendation, “Parameter values for uhdtv systems for production and international programme exchange, itu-r bt. 2020,” *International Telecommunications Union*, 2012.
- [105] W. Reichardt, “Evaluation of optical motion information by movement detectors,” *Journal of Comparative Physiology A*, vol. 161, no. 4, pp. 533–547, 1987.
- [106] E. Reinhard, “Parameter estimation for photographic tone reproduction,” *Journal of Graphics Tools*, vol. 7, no. 1, pp. 45–51, 2002.
- [107] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High Dynamic Range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

- [108] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, “Photographic tone reproduction for digital images,” *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 267–276, Jul. 2002. [Online]. Available: <http://doi.acm.org/10.1145/566654.566575>
- [109] N. Riche, M. Duvinage, M. Mancas, B. Gosselin, and T. Dutoit, “Saliency and human fixations: State-of-the-art and study of comparison metrics,” in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 1153–1160.
- [110] T. Richter, “Backwards compatible coding of high dynamic range images with JPEG,” in *Data Compression Conference (DCC), 2013*. IEEE, 2013, pp. 153–160.
- [111] —, “On the standardization of the JPEG XT image compression,” in *Picture Coding Symposium (PCS), 2013*. IEEE, 2013, pp. 37–40.
- [112] —, “On the integer coding profile of JPEG XT,” in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 9217, 2014, p. 19.
- [113] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable h. 264/MPEG4-AVC extension,” in *2006 IEEE International Conference on Image Processing*. IEEE, 2006, pp. 161–164.
- [114] F. Y. Shih and Y. Yuan, “A wavelet-based encoding algorithm for High Dynamic Range images.” *Open Signal Processing Journal*, vol. 3, 2010.
- [115] M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta, “A standard default color space for the internet-srgb,” *Microsoft and Hewlett-Packard Joint Report*, 1996.
- [116] M. Sugawara, S.-Y. Choi, and D. Wood, “Ultra-high-definition television (rec. itu-r bt. 2020): A generational leap in the evolution of television

- [standards in a nutshell],” *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 170–174, 2014.
- [117] N. Sugiyama, H. Kaida, X. Xue, T. Jinno, N. Adami, and M. Okuda, “Hdr image compression using optimized tone mapping model,” in *ICASSP 2009. IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 1001–1004.
- [118] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (hevc) standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [119] W. Sweldens, “The lifting scheme: A custom-design construction of biorthogonal wavelets,” *Applied and Computational Harmonic Analysis*, vol. 3, no. 2, pp. 186–200, 1996.
- [120] —, “The lifting scheme: A construction of second generation wavelets,” *SIAM journal on Mathematical Analysis*, vol. 29, no. 2, pp. 511–546, 1998.
- [121] K. H. Talukder and K. Harada, “Haar wavelet based approach for image compression and quality assessment of compressed image,” *arXiv preprint arXiv:1010.4084*, 2010.
- [122] D. Taubman and M. Marcellin, *JPEG2000 image compression fundamentals, standards and practice: image compression fundamentals, standards and practice*. Springer Science & Business Media, 2012, vol. 642.
- [123] D. Touzé, S. Lasserre, Y. Olivier, R. Boitard, and E. François, “Hdr video coding based on local ldr quantization,” in *HDRi2014-Second International Conference and SME Workshop on HDR imaging*, 2014.
- [124] D. Tsai, M. Flagg, A. Nakazawa, and J. M. Rehg, “Motion coherent

- tracking using multi-label mrf optimization,” *International journal of Computer Vision*, vol. 100, no. 2, pp. 190–202, 2012.
- [125] J. Tumblin and H. Rushmeier, “Tone reproduction for realistic images,” *IEEE Computer Graphics and Applications*, vol. 13, no. 6, pp. 42–48, 1993.
- [126] S. Vasireddy, “Hdr video compression using h. 265 (hevc) main 10 profile,” Ph.D. dissertation, UNIVERSITY OF TEXAS AT ARLINGTON, 2016.
- [127] G. K. Wallace, “The JPEG still picture compression standard,” *IEEE Transactions on Consumer Electronics*, vol. 38, no. 1, pp. xviii–xxxiv, 1992.
- [128] Z. Wang, L. Lu, and A. C. Bovik, “Foveation scalable video coding with automatic fixation selection,” *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 243–254, 2003.
- [129] G. Ward, “Real pixels,” *Graphics Gems II*, pp. 80–83, 1991.
- [130] —, “High Dynamic Range image examples in rgbe formats,” <http://www.anywhere.com/gward/hdrenc/pages/originals.html>, Sept 2016.
- [131] G. Ward and M. Simmons, “JPEG-HDR: A backwards-compatible, High Dynamic Range extension to JPEG,” in *ACM SIGGRAPH 2006 Courses*. ACM, 2006, p. 3.
- [132] A. Watanabe, T. Mori, S. Nagata, and K. Hiwatashi, “Spatial sine-wave responses of the Human Visual System,” *Vision Research*, vol. 8, no. 9, pp. 1245–1263, 1968.
- [133] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, “Overview of the h. 264/avc video coding standard,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

- [134] G. Wyszecki and W. S. Stiles, *Color science*. Wiley New York, 1982, vol. 8.
- [135] R. Xu, S. N. Pattanaik, and C. E. Hughes, “High-Dynamic-Range Still-Image Encoding in JPEG 2000,” *IEEE Comput. Graph. Appl.*, vol. 25, no. 6, pp. 57–64, Nov. 2005. [Online]. Available: <http://dx.doi.org/10.1109/MCG.2005.133>
- [136] C. Yee and D. Walther, “Motion detection for bottom-up visual attention,” tech. rep., SURF/CNS, California Institute of Technology, Tech. Rep., 2002.
- [137] A. Yoshida, V. Blanz, K. Myszkowski, and H.-P. Seidel, “Perceptual evaluation of tone mapping operators with real-world scenes,” in *Electronic Imaging 2005*. International Society for Optics and Photonics, 2005, pp. 192–203.
- [138] Y. Zhai and M. Shah, “Visual attention detection in video sequences using spatiotemporal cues,” in *Proceedings of the 14th ACM international conference on Multimedia*. ACM, 2006, pp. 815–824.
- [139] Y. Zhang, D. Agrafiotis, M. Naccari, M. Mrak, and D. R. Bull, “Visual masking phenomena with high dynamic range content,” in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 2284–2288.
- [140] Y. Zhang, M. Naccari, D. Agrafiotis, M. Mrak, and D. R. Bull, “High dynamic range video compression exploiting luminance masking,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 5, pp. 950–964, May 2016.
- [141] —, “High dynamic range video compression by intensity dependent spatial quantization in hevc,” in *Picture Coding Symposium (PCS), 2013*. IEEE, 2013, pp. 353–356.

- [142] Y. Zhang, E. Reinhard, and D. Bull, “Perception-based high dynamic range video compression with optimal bit-depth transformation,” in *2011 18th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 1321–1324.
- [143] S.-h. Zhong, Y. Liu, T.-Y. Ng, and Y. Liu, “Perception-oriented video saliency detection via spatio-temporal attention analysis,” *Neurocomputing*, 2016.
- [144] S.-h. Zhong, Y. Liu, F. Ren, J. Zhang, and T. Ren, “Video saliency detection via Dynamic consistent spatio-temporal attention modelling.” in *AAAI*, 2013.

Appendix A

Data-set

The thumbnails of the data-set are presented following. For compression in chapter 3 and 4:

RGBE images:



Figure A1: RGBE images in data-set.



Figure A2: RGBE images in data-set.

EXR images:

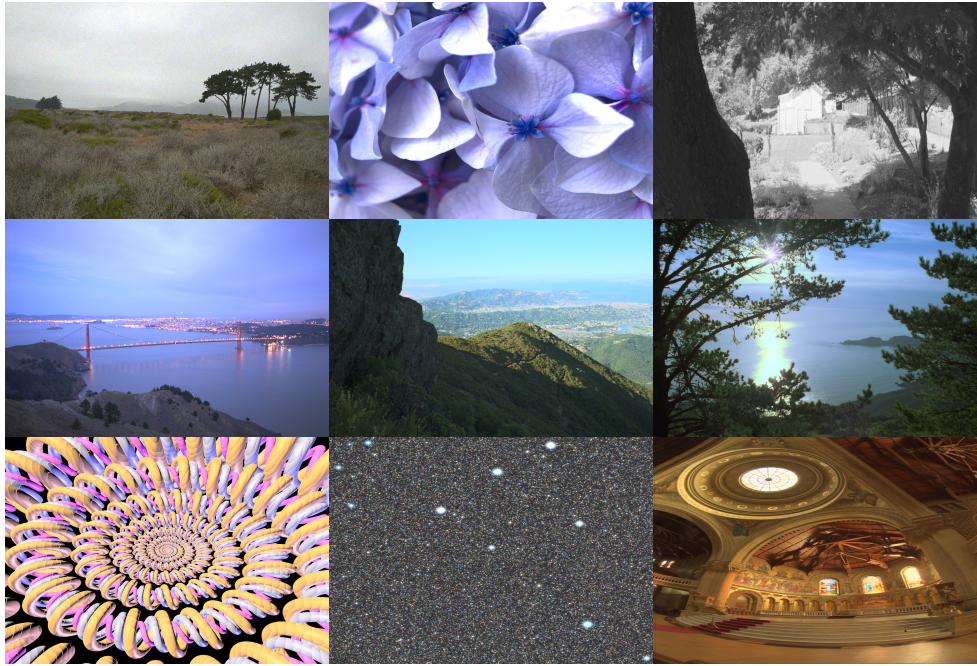


Figure A3: EXR images in data-set.

PFM images:

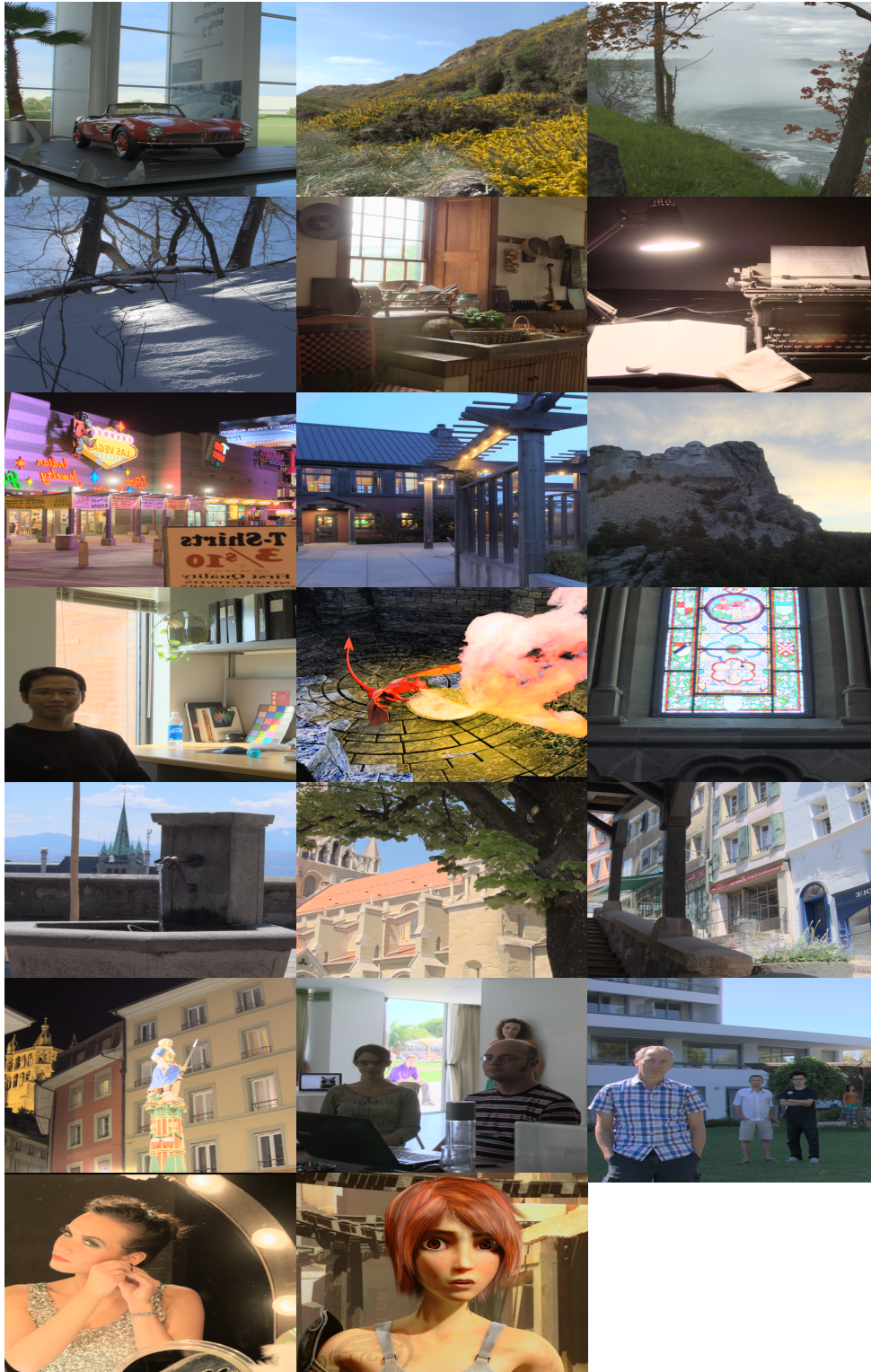


Figure A4: PFM images in data-set.

Data-set for saliency estimation in the Chapter 5:

RGBE images:

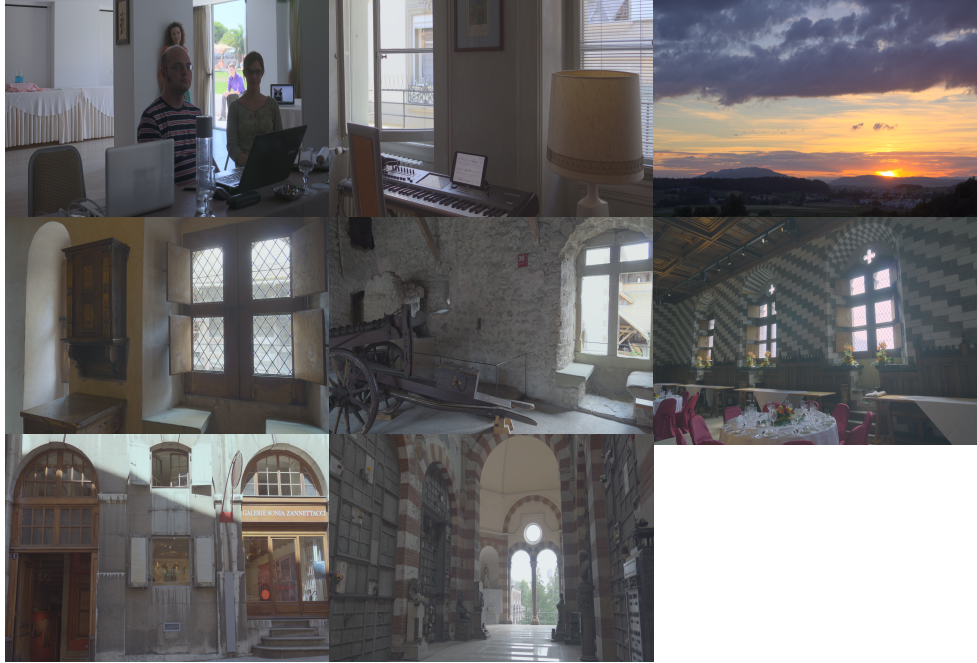


Figure A5: Saliency estimation data-set.

EXR Sequences:



Figure A6: EXR Sequences in data-set.

Appendix B

Full RD Results of Chapter 3

The RD results of Chapter 3 are listed: **RGBE image results:**

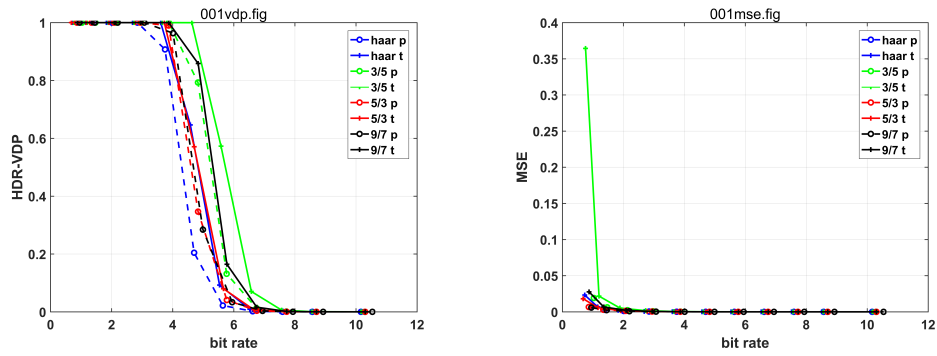


Figure A7: RGBE 001.

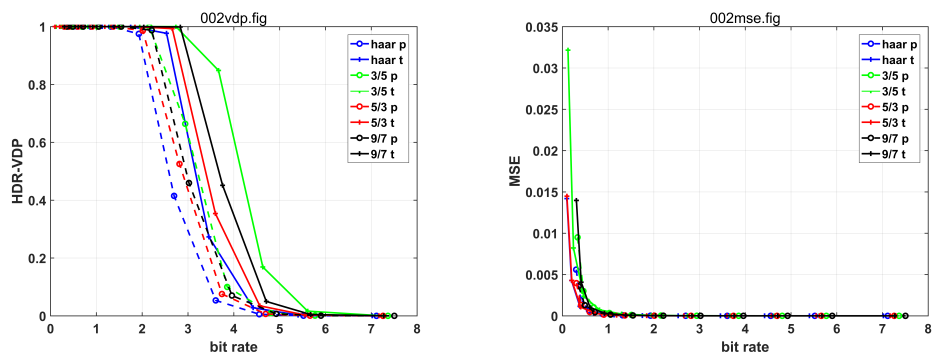


Figure A8: RGBE 002.

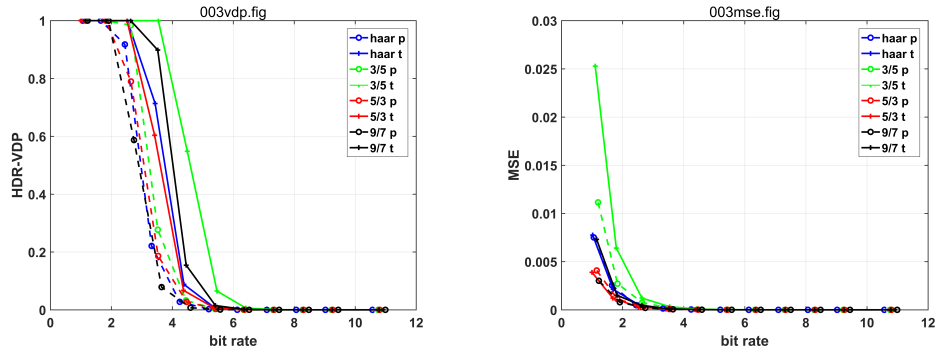


Figure A9: RGBE 003.

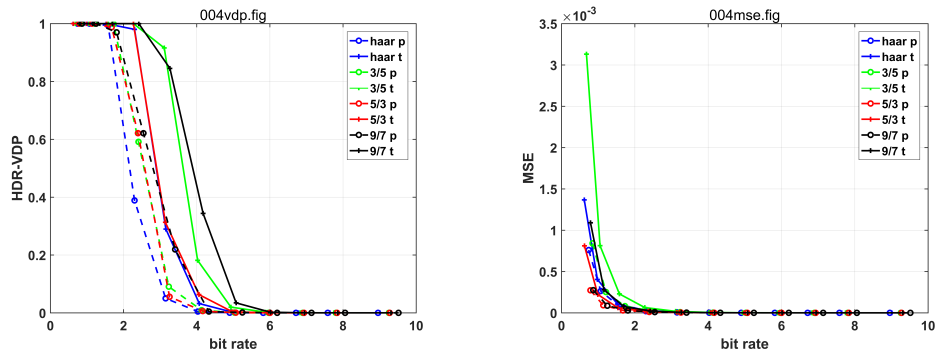


Figure A10: RGBE 004.

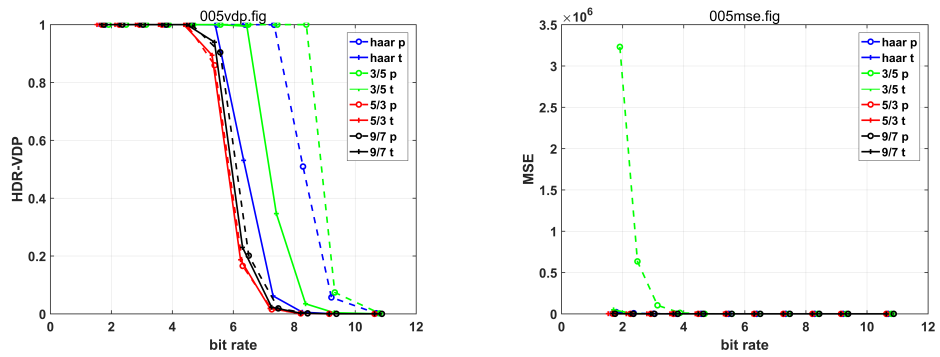


Figure A11: RGBE 005.

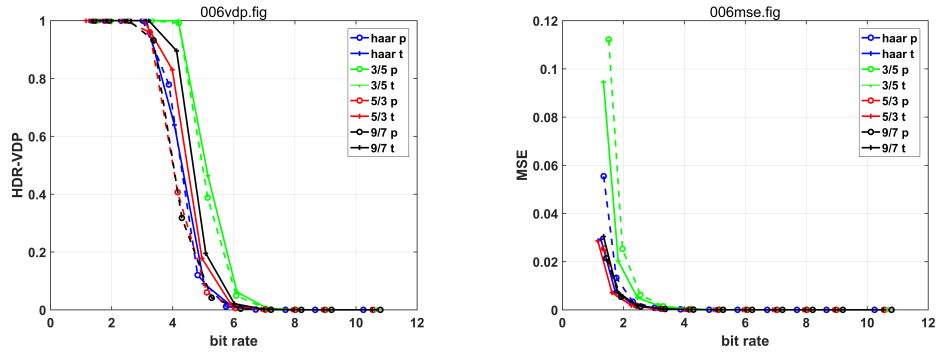


Figure A12: RGBE 006.

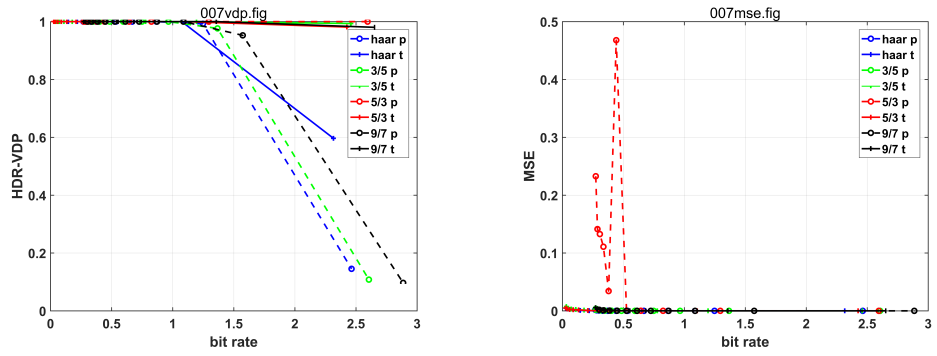


Figure A13: RGBE 007.

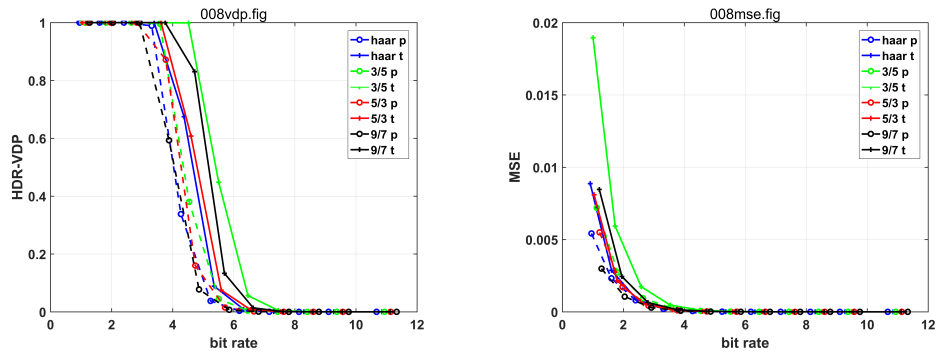


Figure A14: RGBE 008.

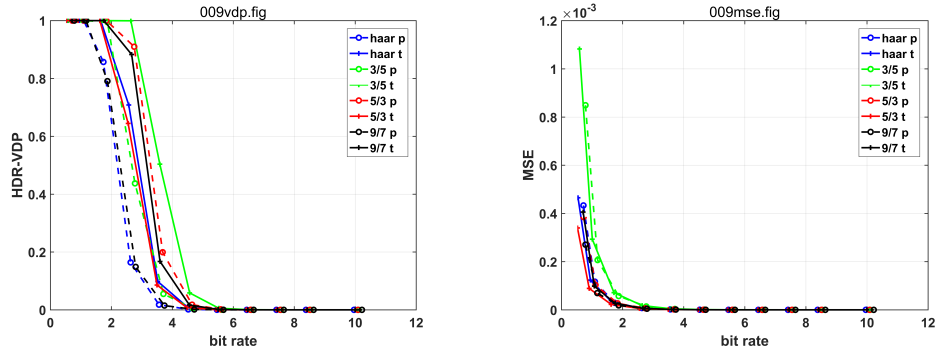


Figure A15: RGBE 009.

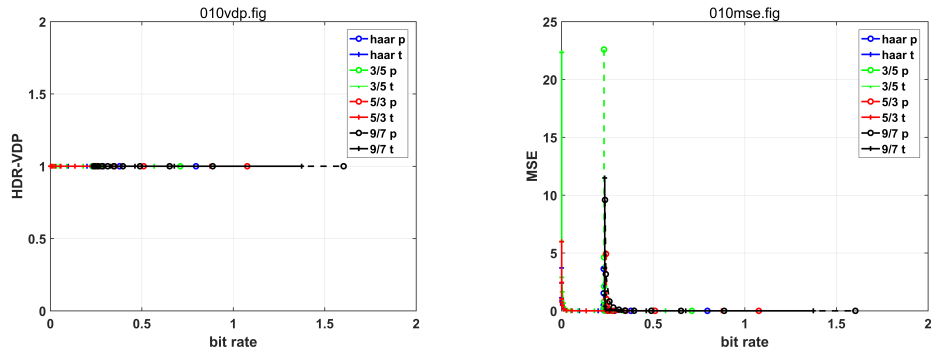


Figure A16: RGBE 010.

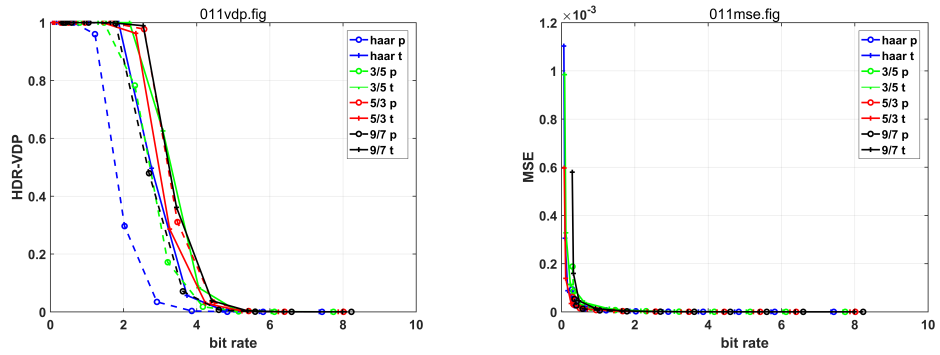


Figure A17: RGBE 011.

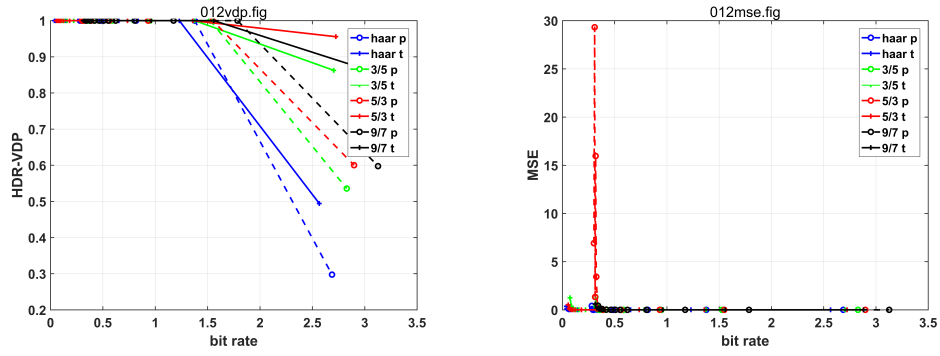


Figure A18: RGBE 012.

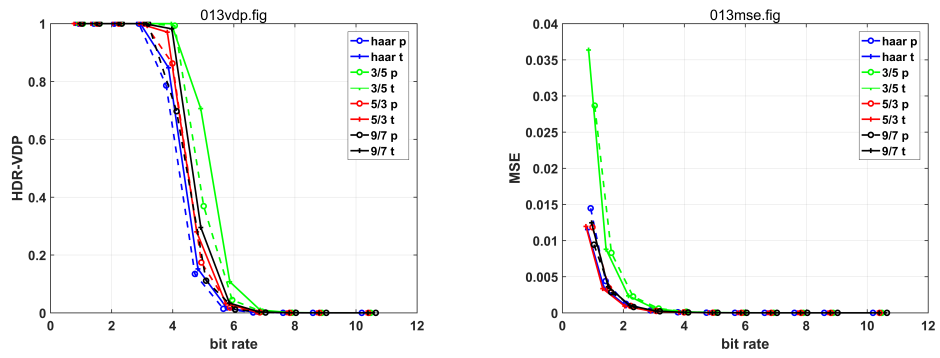


Figure A19: RGBE 013.

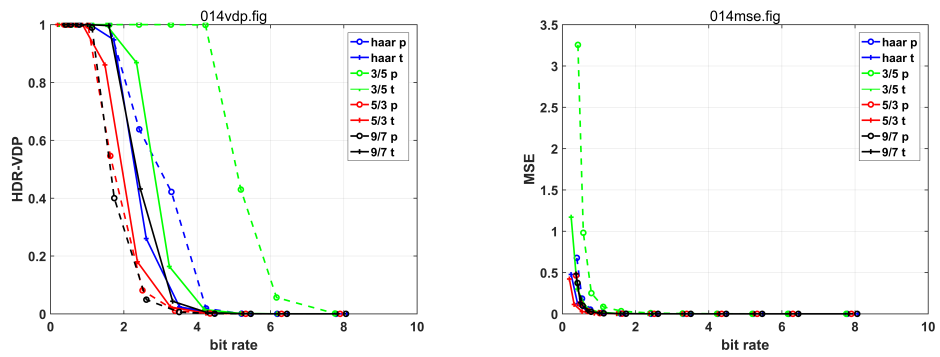


Figure A20: RGBE 014.

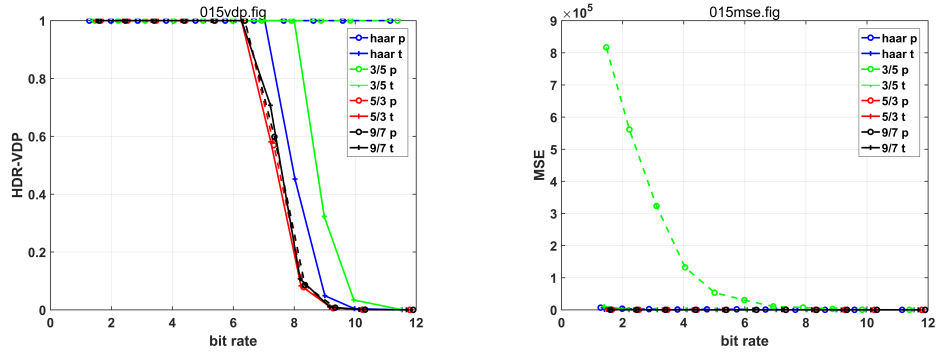


Figure A21: RGBE 015.

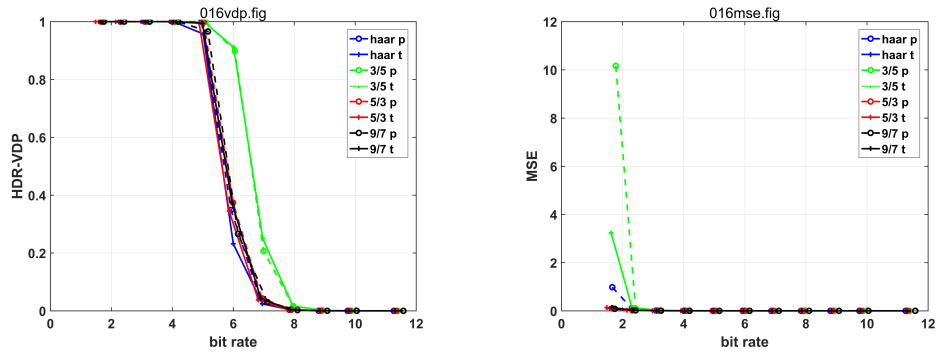


Figure A22: RGBE 016.

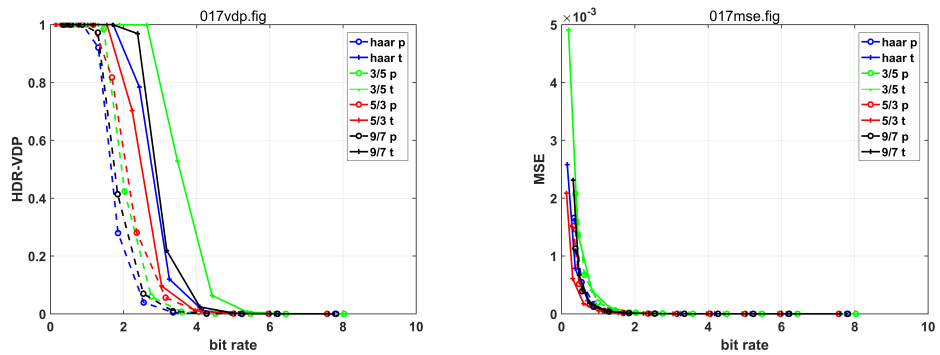


Figure A23: RGBE 017.

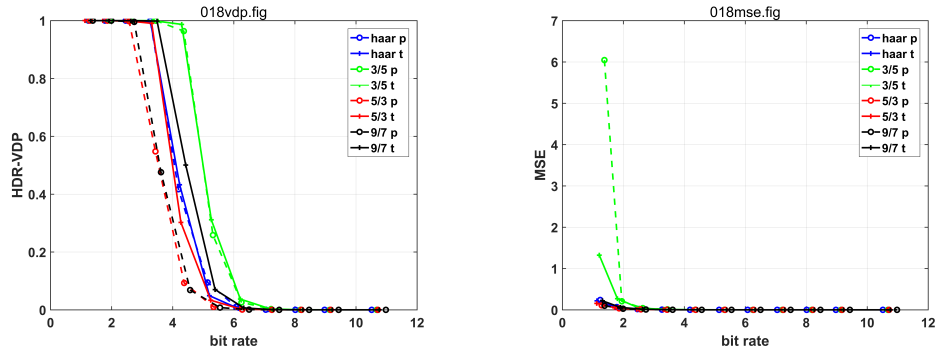


Figure A24: RGBE 018.

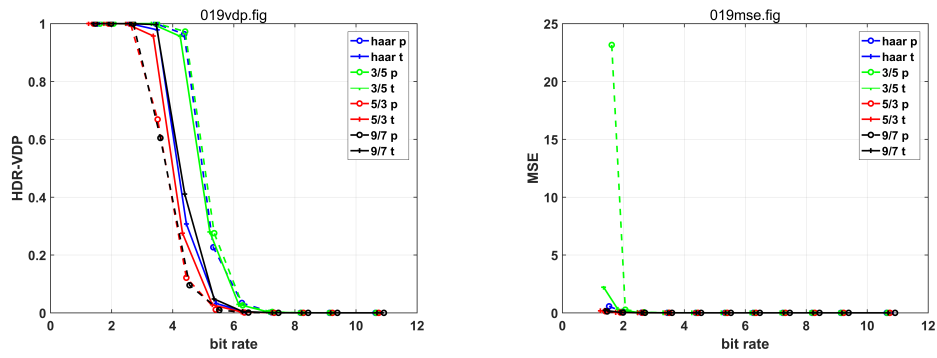


Figure A25: RGBE 019.

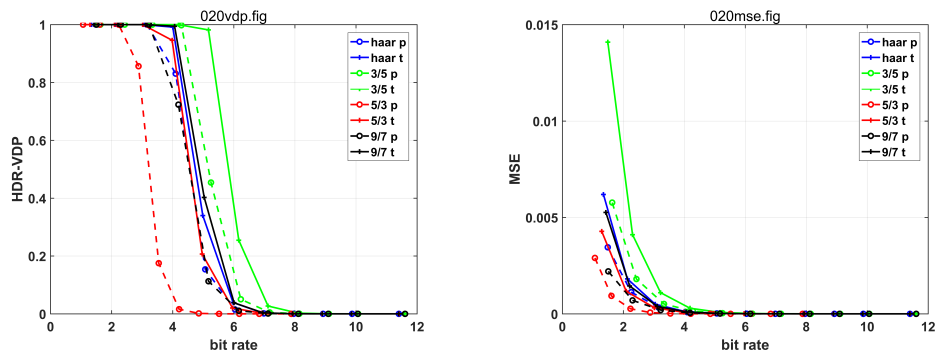


Figure A26: RGBE 020.

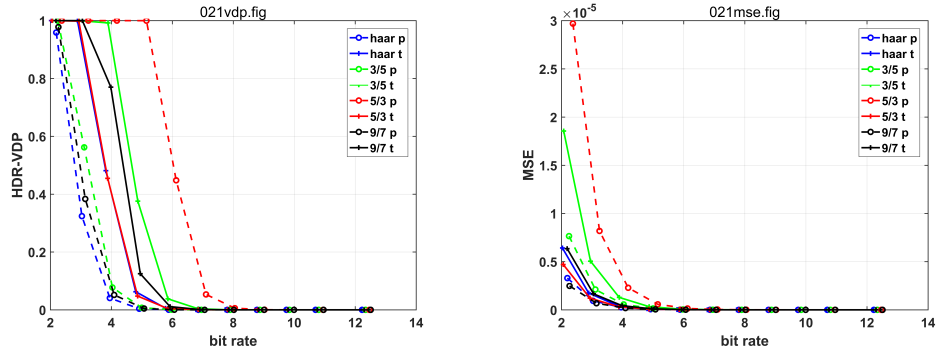


Figure A27: RGBE 021.

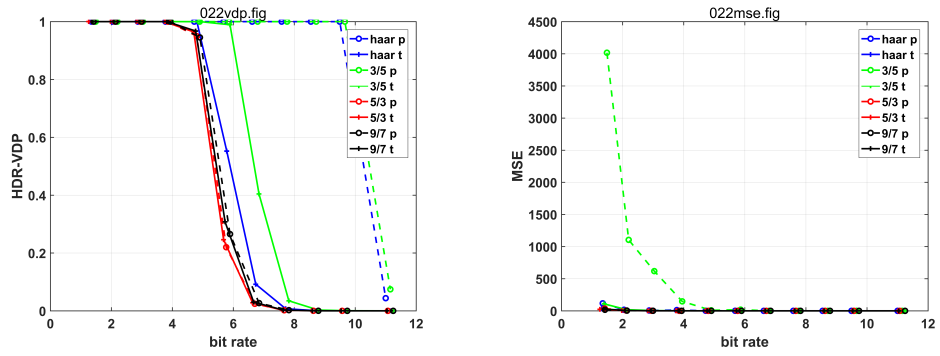


Figure A28: RGBE 022.

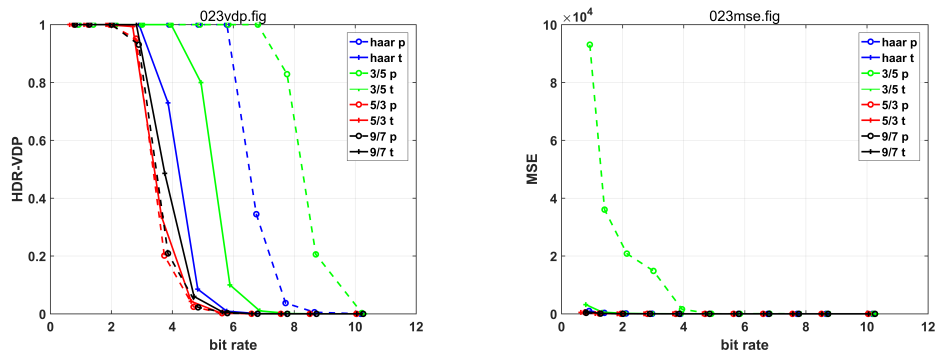


Figure A29: RGBE 023.

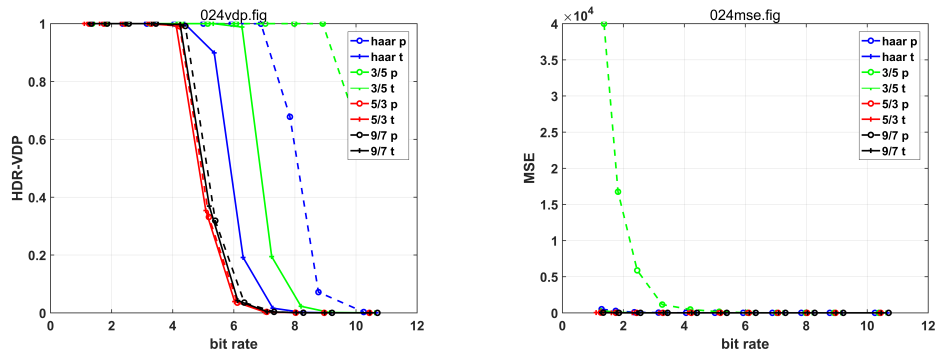


Figure A30: RGBE 024.

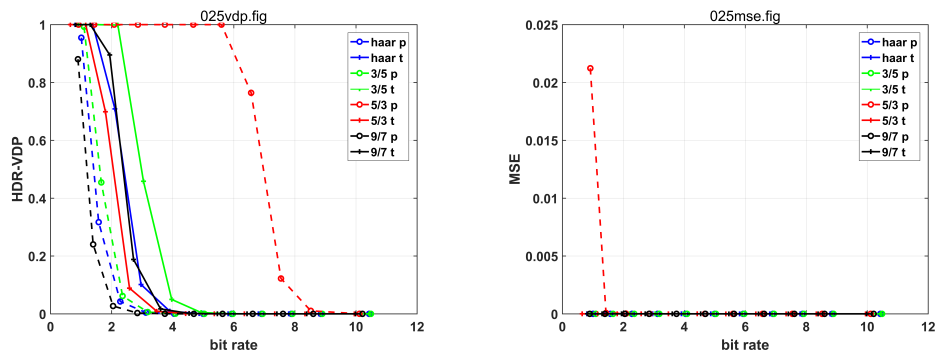


Figure A31: RGBE 025.

EXR image results:

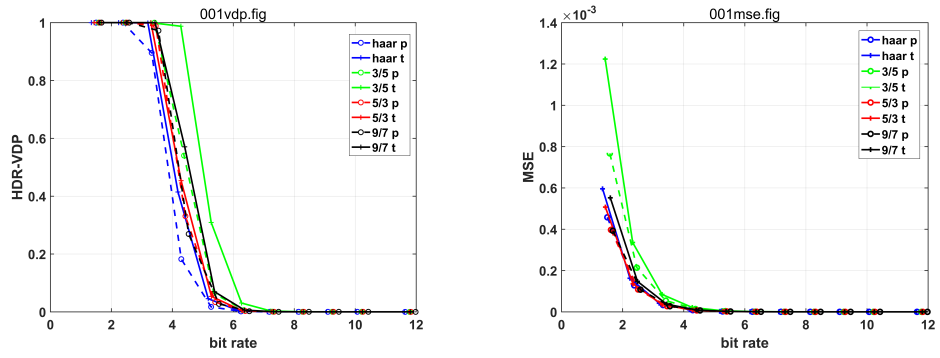


Figure A32: EXR 001.

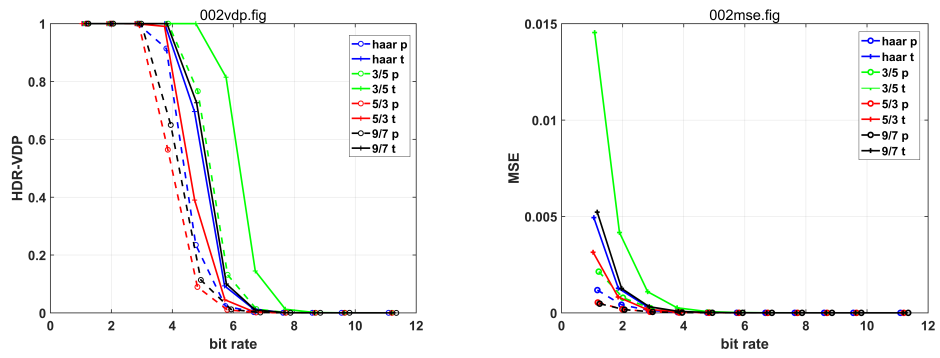


Figure A33: EXR 002.

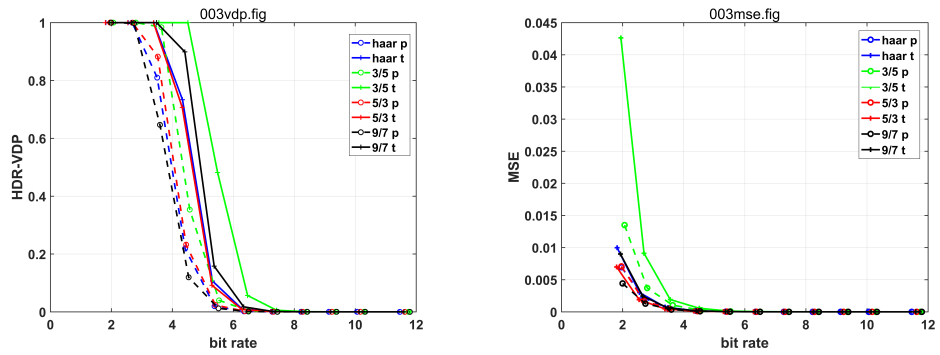


Figure A34: EXR 003.

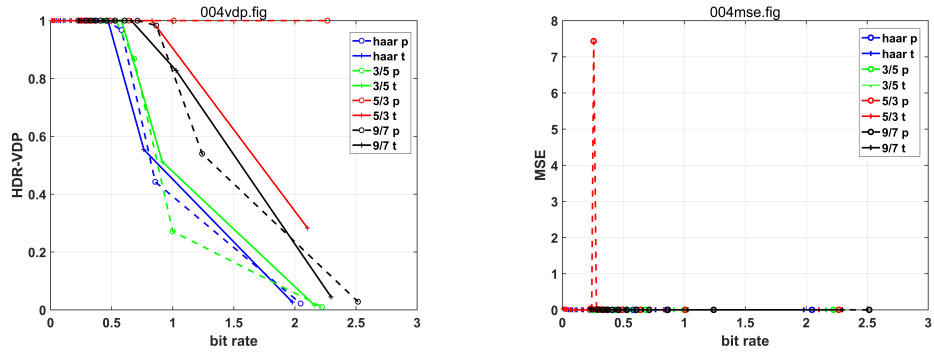


Figure A35: EXR 004.

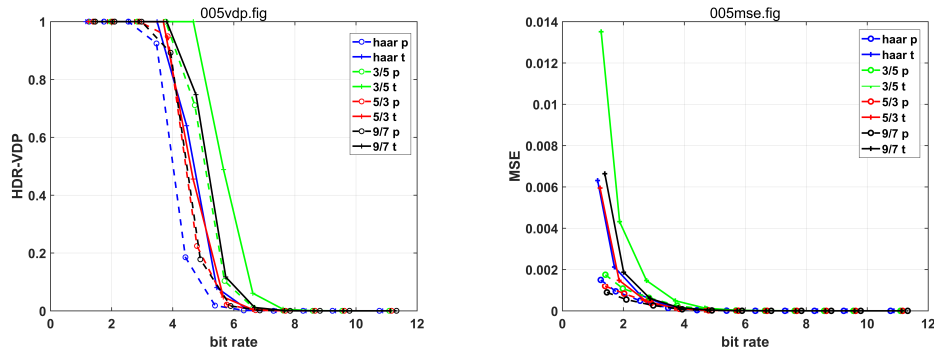


Figure A36: EXR 005.

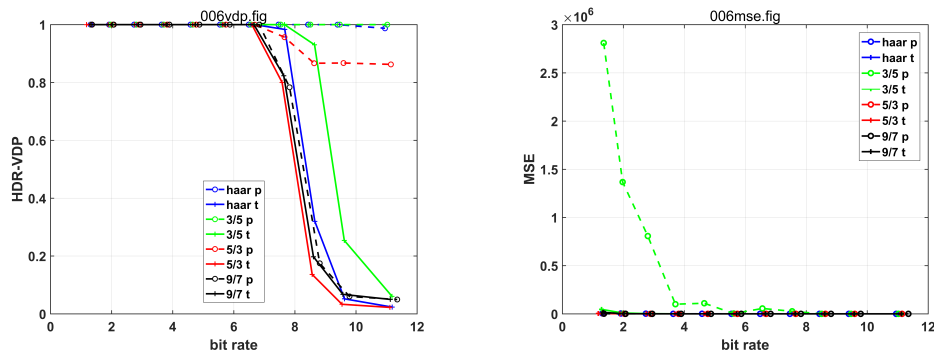


Figure A37: EXR 006.

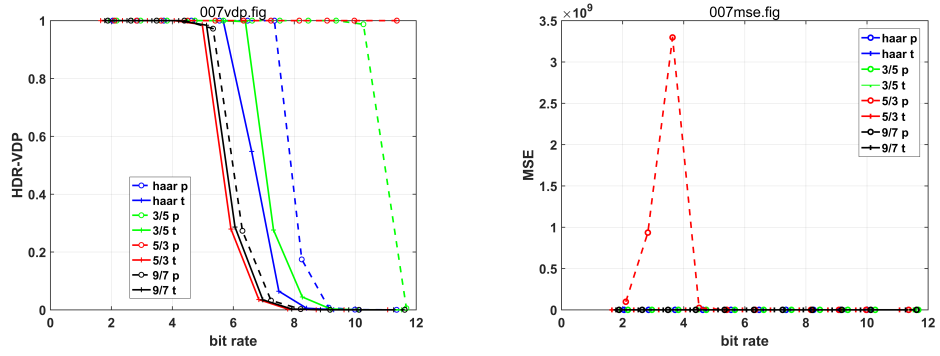


Figure A38: EXR 007.

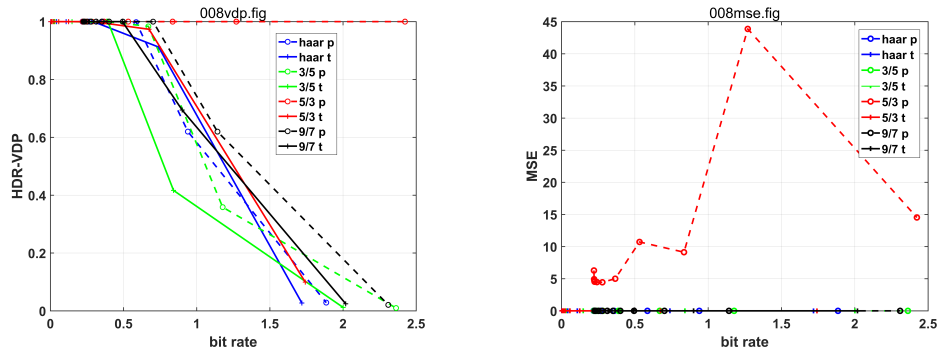


Figure A39: EXR 008.

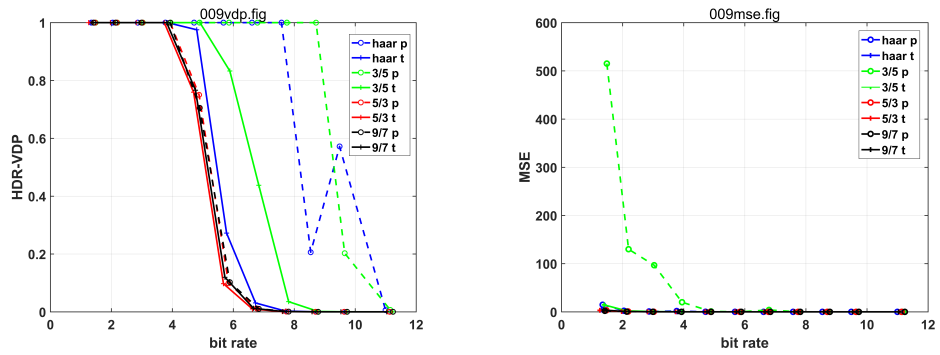


Figure A40: EXR 009.

Appendix C

Full RD Results of the Chapter 4

The RD results of Chapter 4 are listed: **IMAGE results**

RGBE:

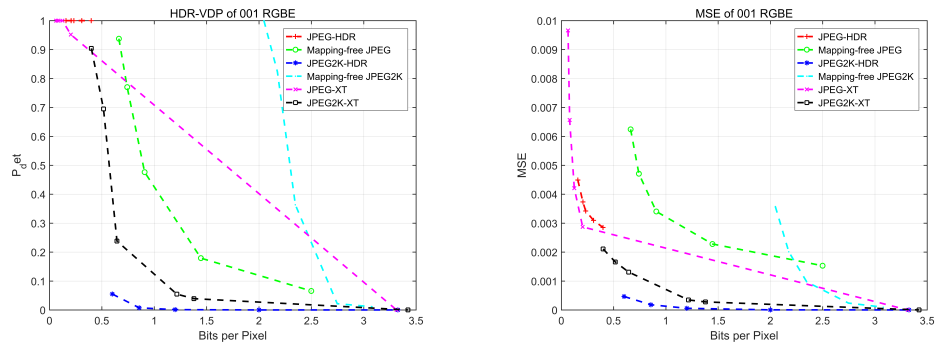


Figure A41: RGBE 001.

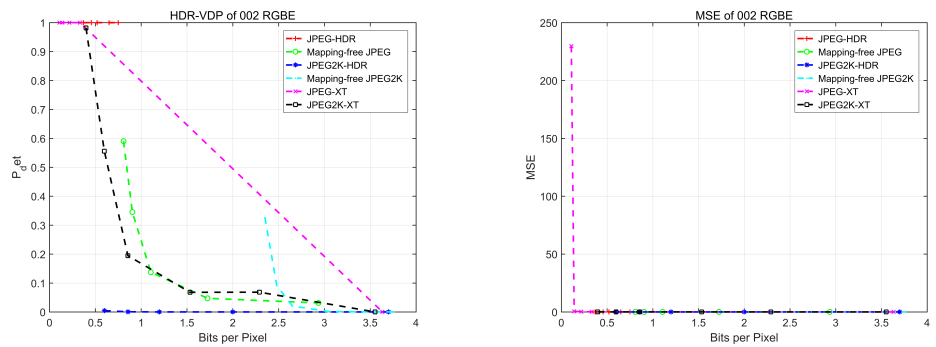


Figure A42: RGBE 002.

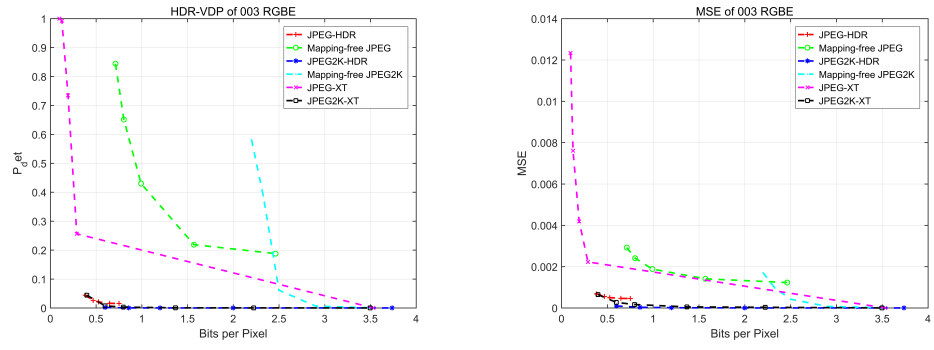


Figure A43: RGBE 003.

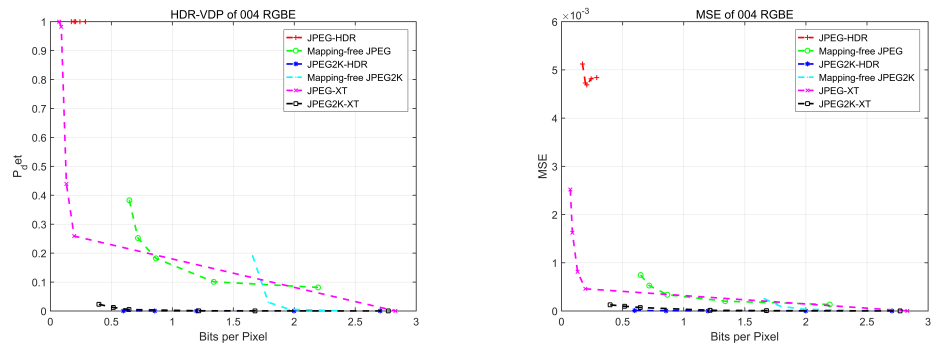


Figure A44: RGBE 004.

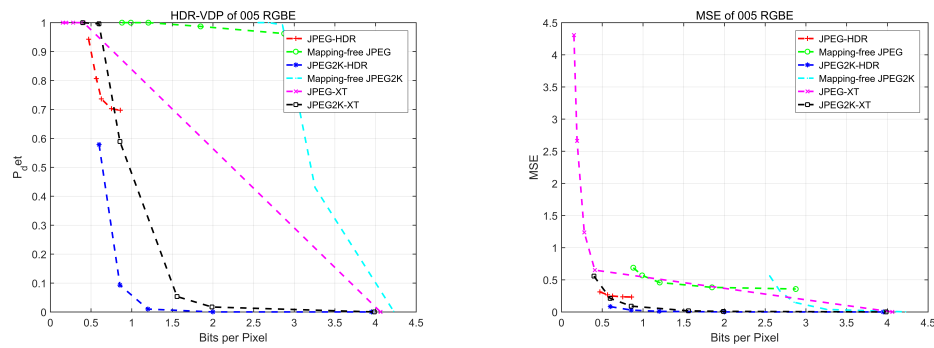


Figure A45: RGBE 005.

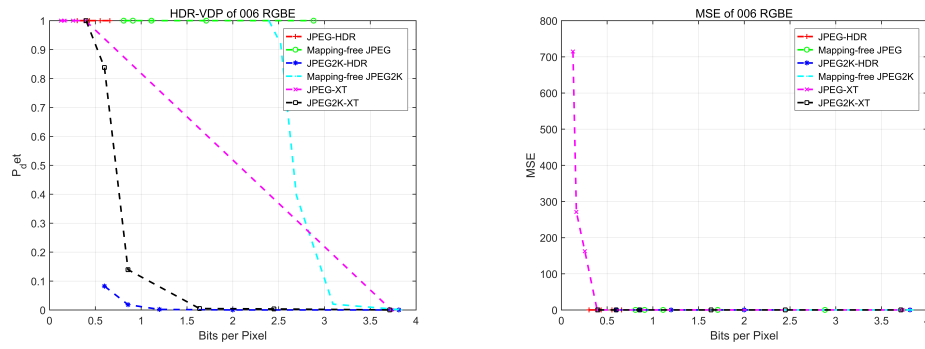


Figure A46: RGBE 006.

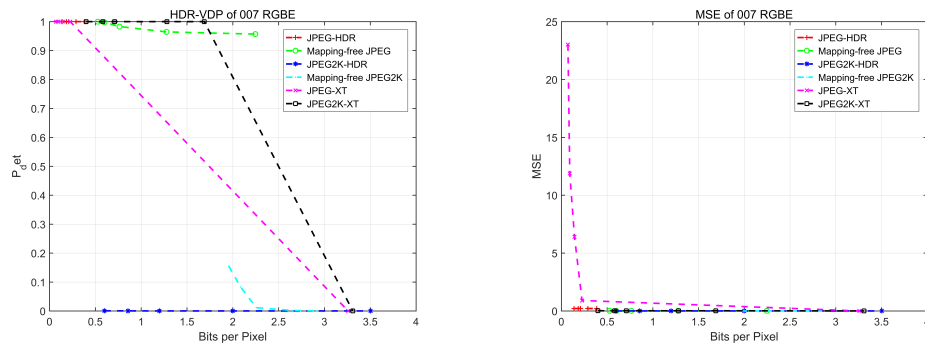


Figure A47: RGBE 007.

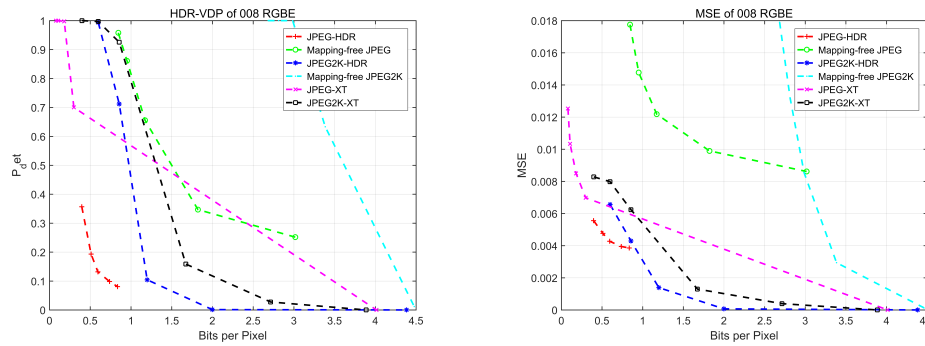


Figure A48: RGBE 008.

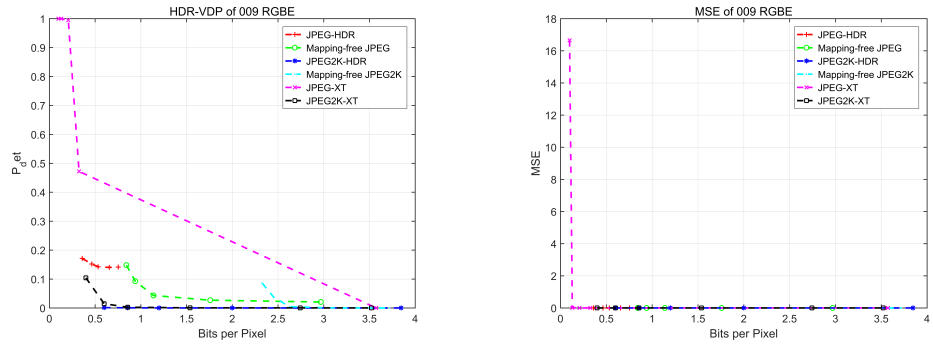


Figure A49: RGBE 009.

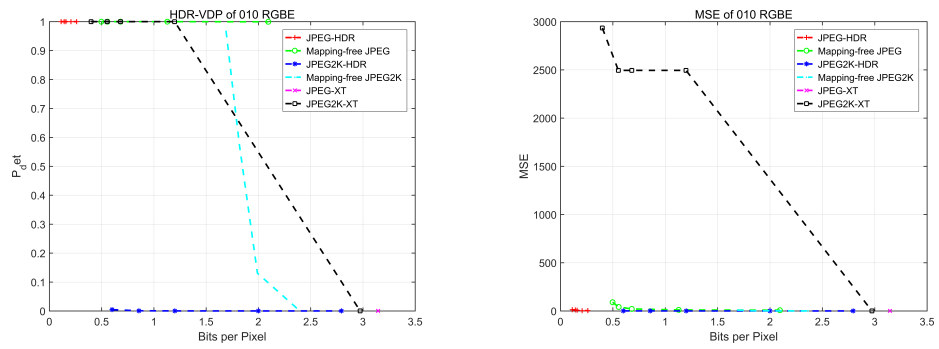


Figure A50: RGBE 010.

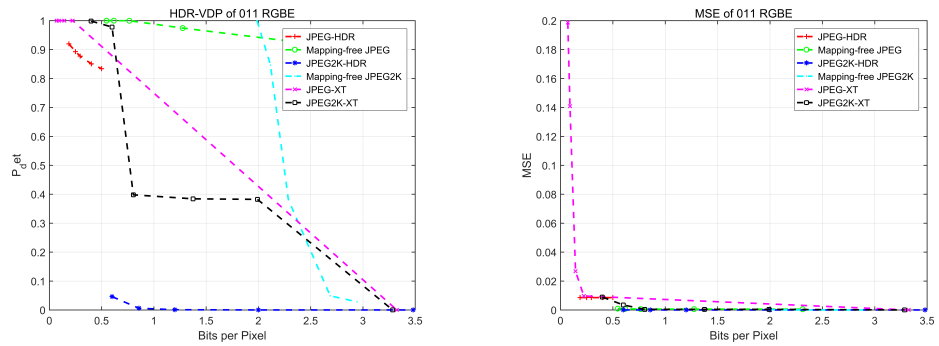


Figure A51: RGBE 011.

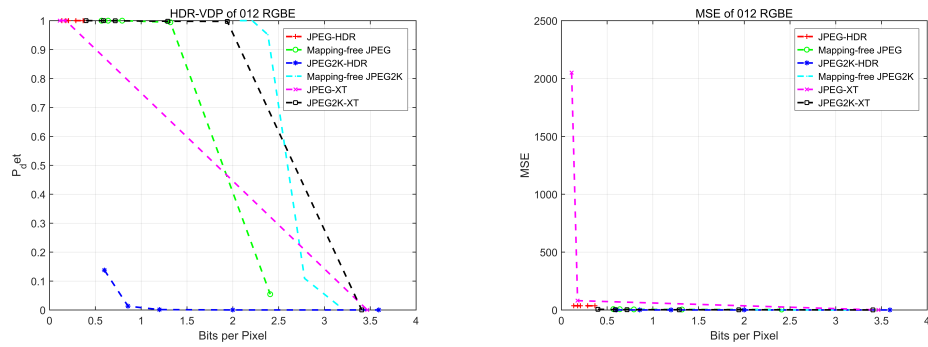


Figure A52: RGBE 012.

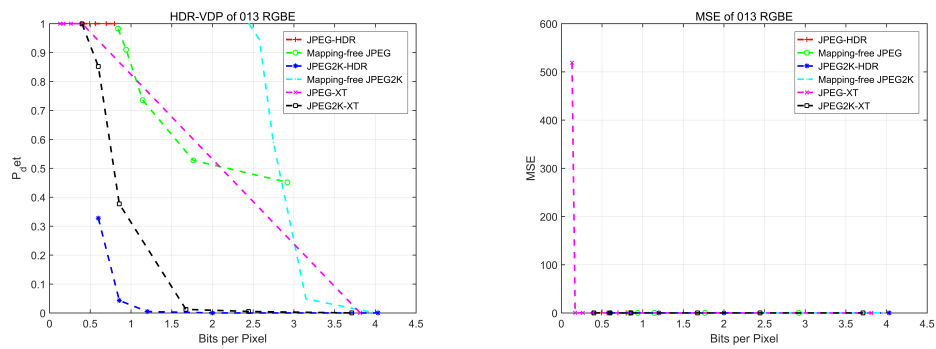


Figure A53: RGBE 013.

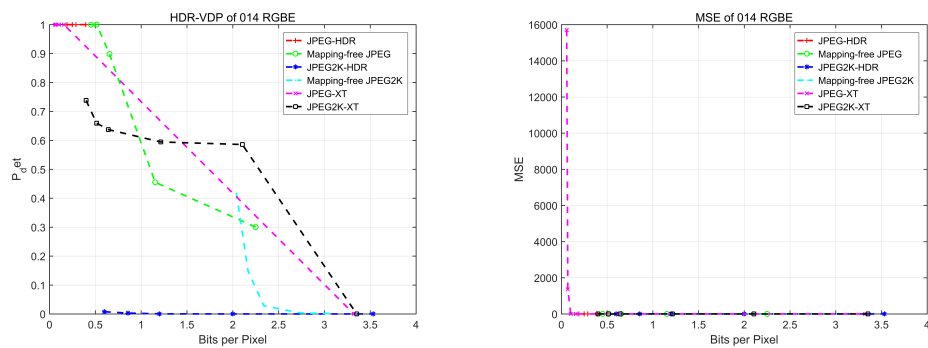


Figure A54: RGBE 014.

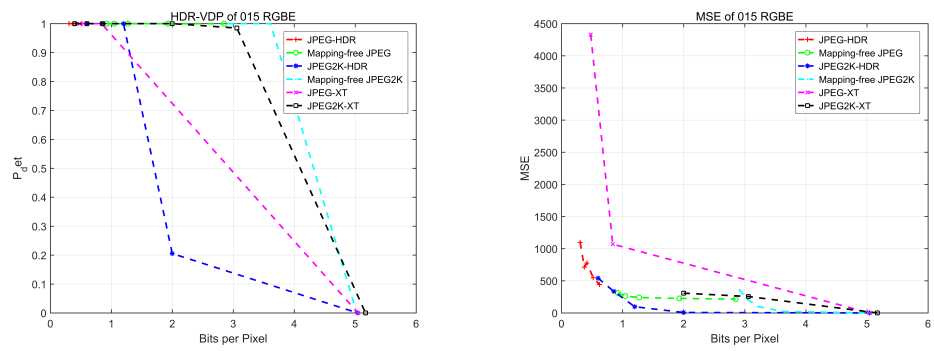


Figure A55: RGBE 015.

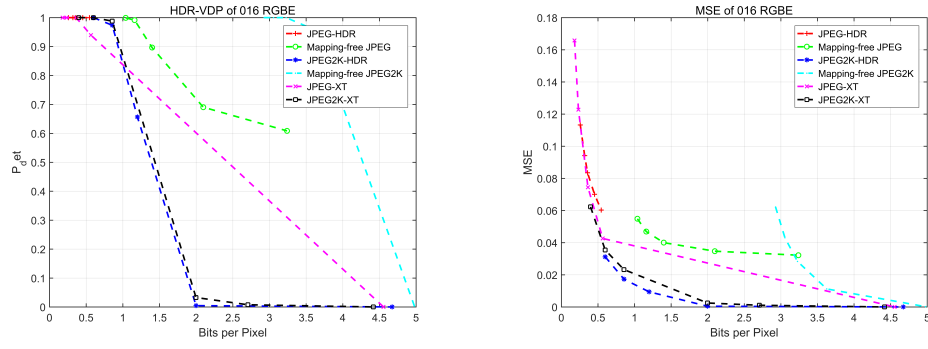


Figure A56: RGBE 016.

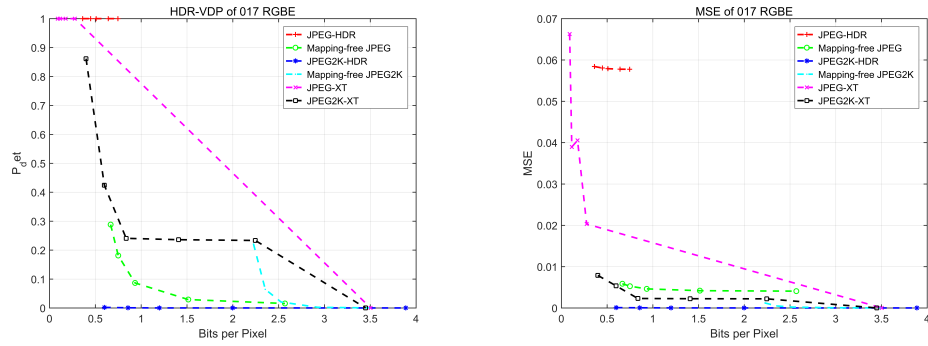


Figure A57: RGBE 017.

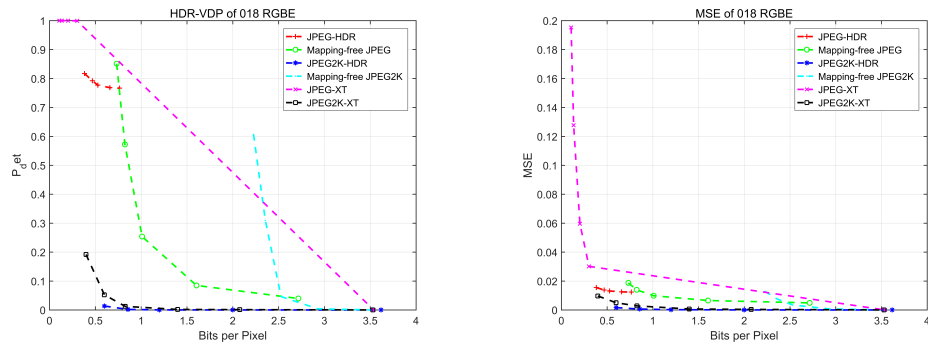


Figure A58: RGBE 018.

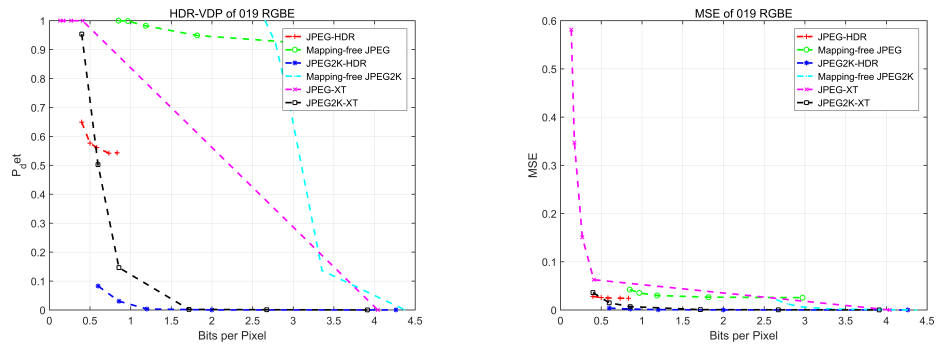


Figure A59: RGBE 019.

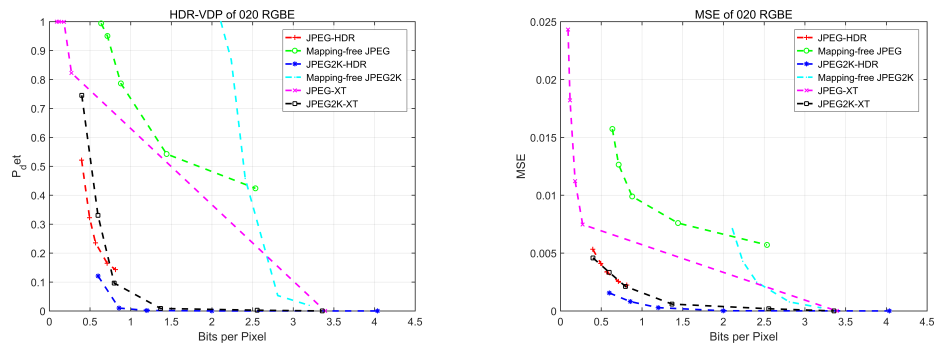


Figure A60: RGBE 020.

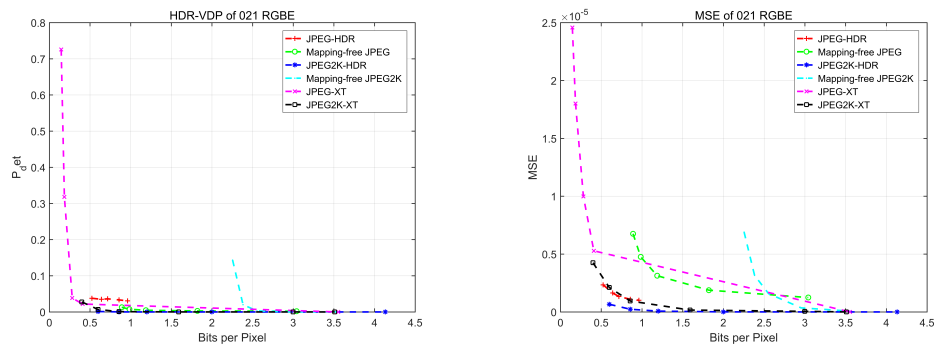


Figure A61: RGBE 021.

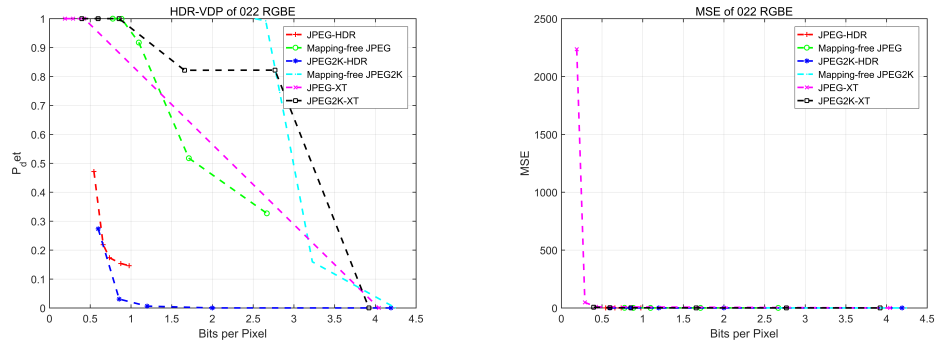


Figure A62: RGBE 022.

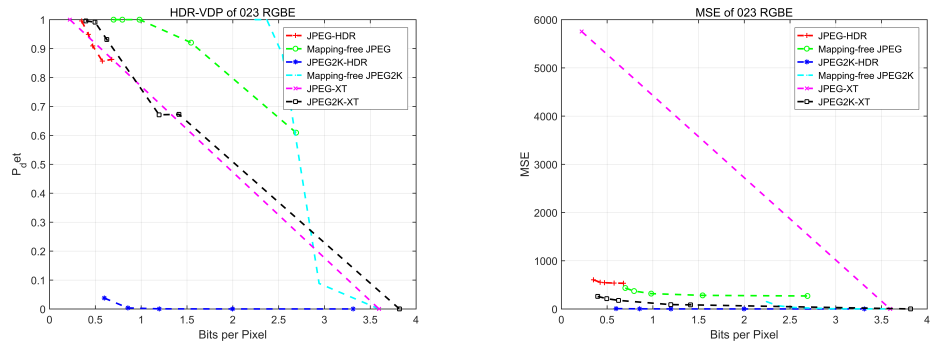


Figure A63: RGBE 023.

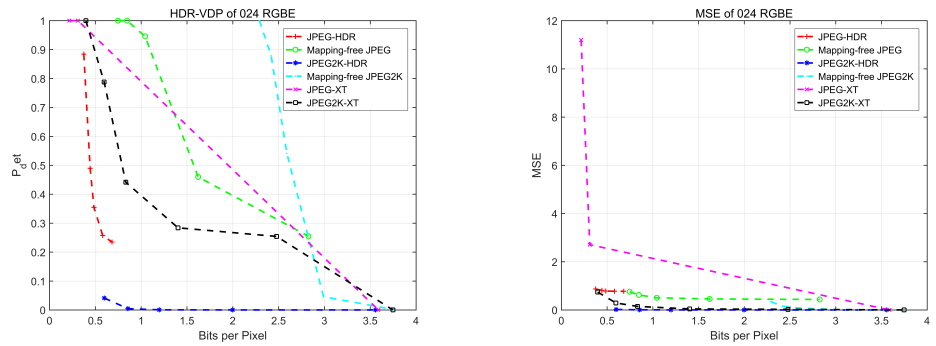


Figure A64: RGBE 024.

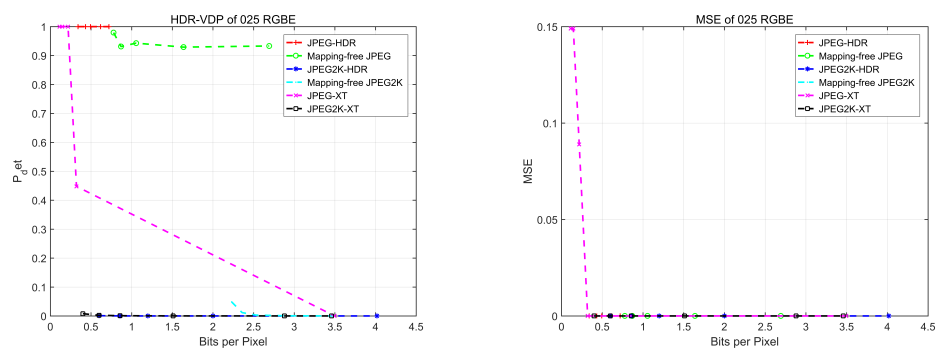


Figure A65: RGBE 025.

EXR:

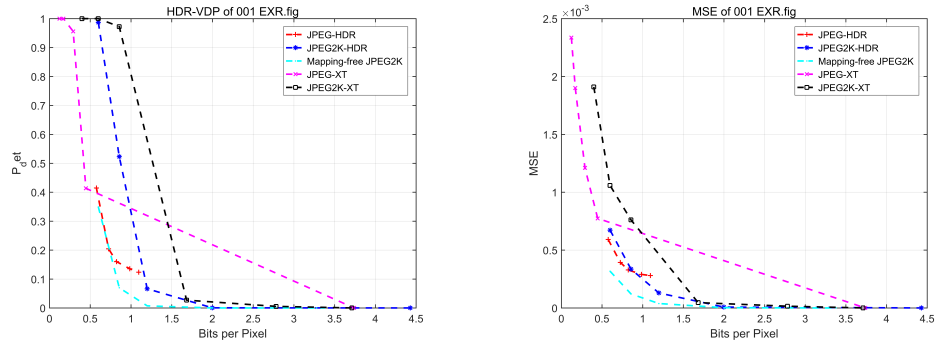


Figure A66: EXR 001.

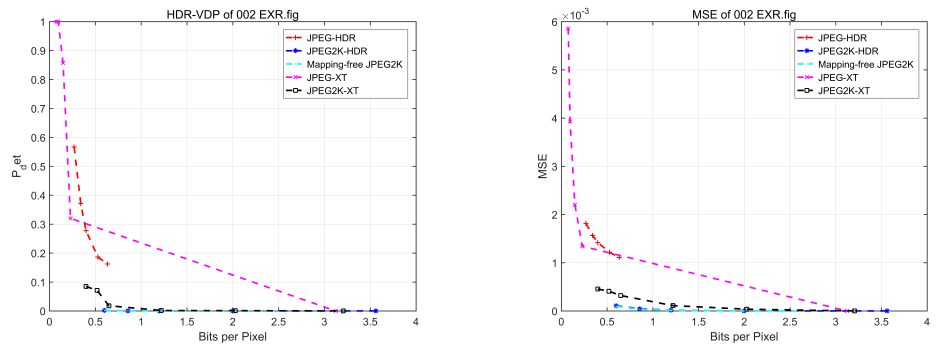


Figure A67: EXR 002.

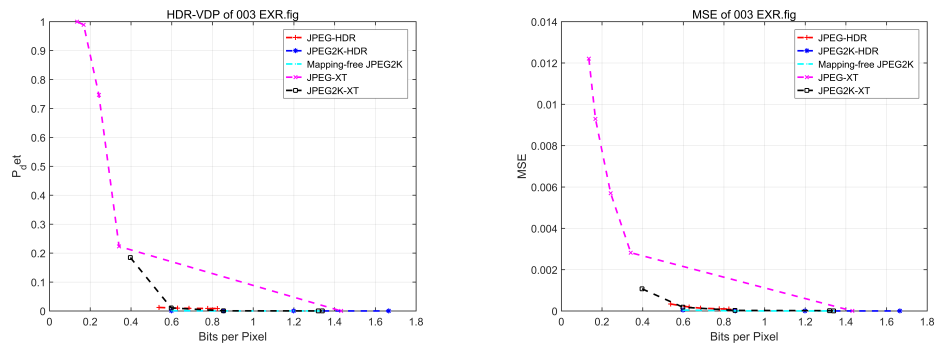


Figure A68: EXR 003.

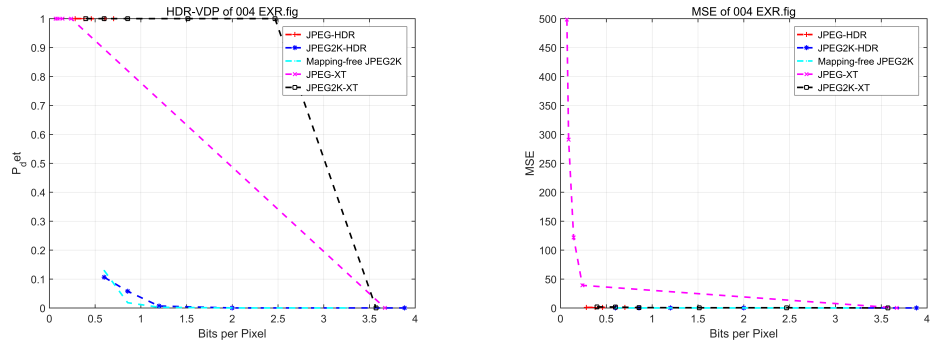


Figure A69: EXR 004.

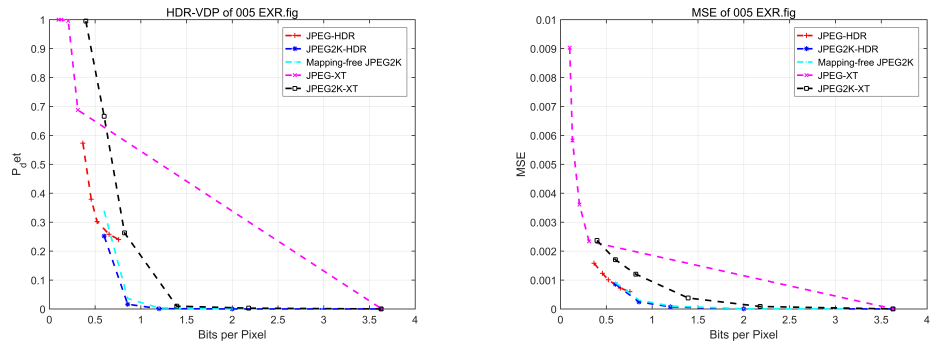


Figure A70: EXR 005.

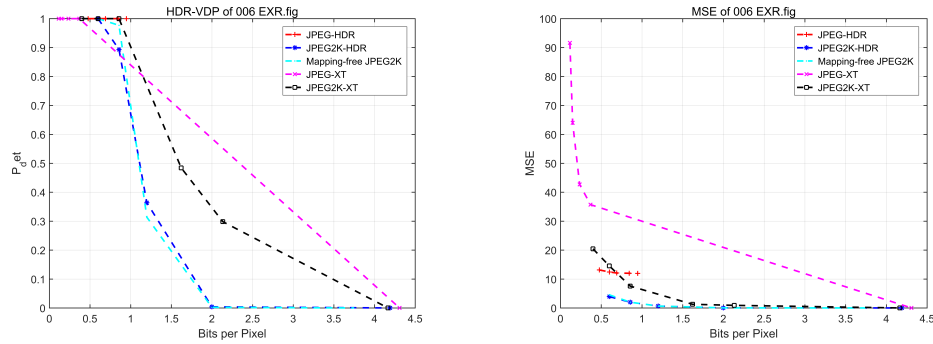


Figure A71: EXR 006.

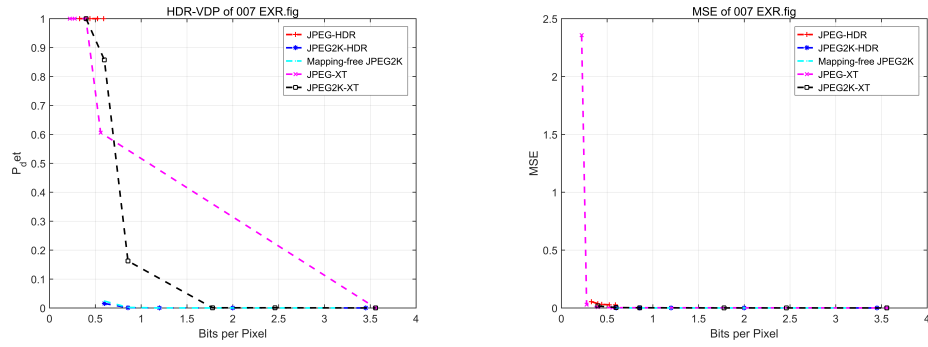


Figure A72: EXR 007.

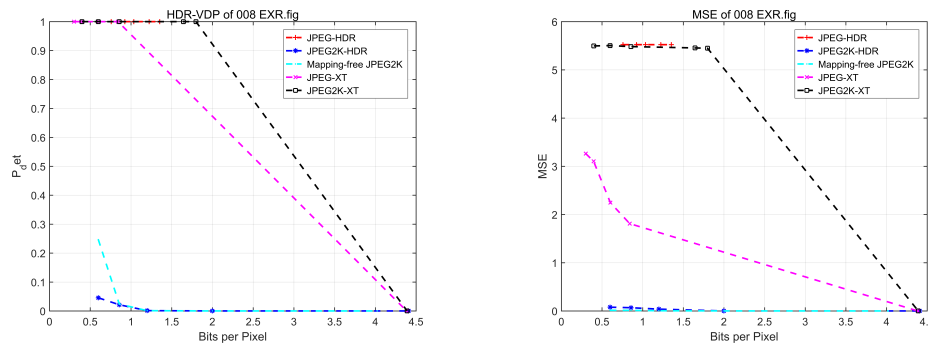


Figure A73: EXR 008.

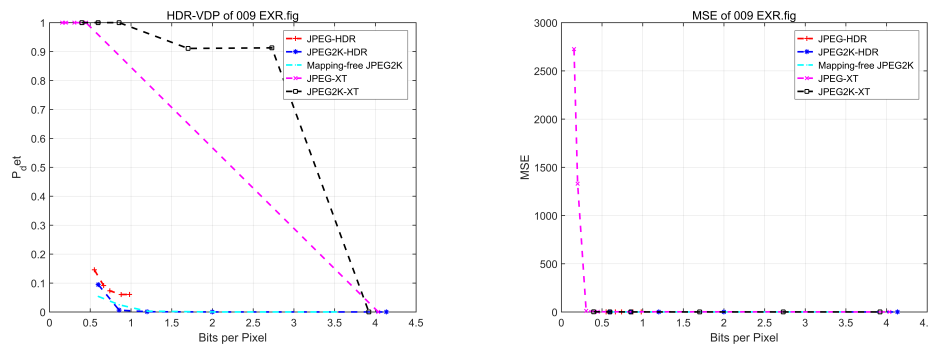


Figure A74: EXR 009.

PFM:

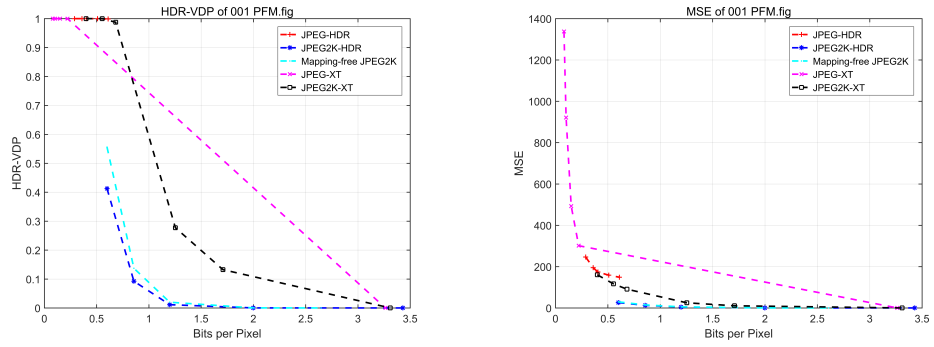


Figure A75: PFM 001.

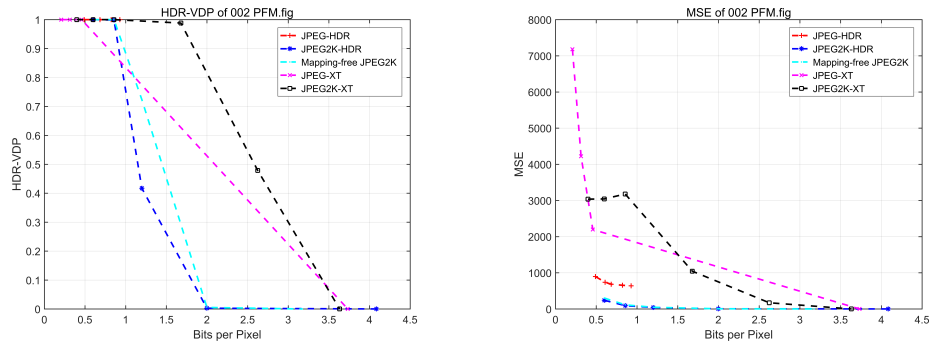


Figure A76: PFM 002.

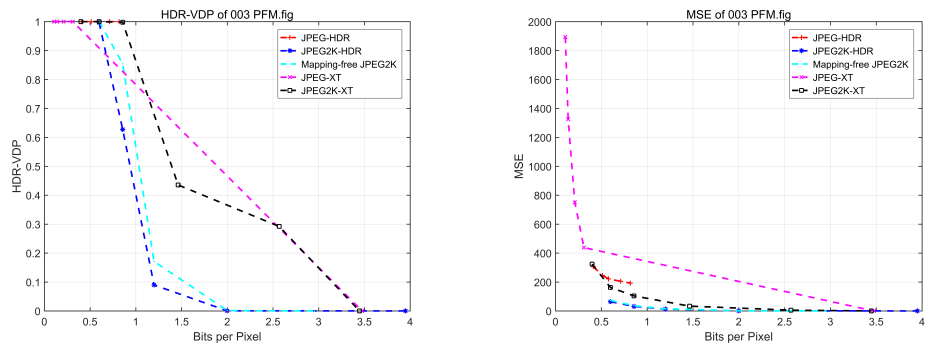


Figure A77: PFM 003.

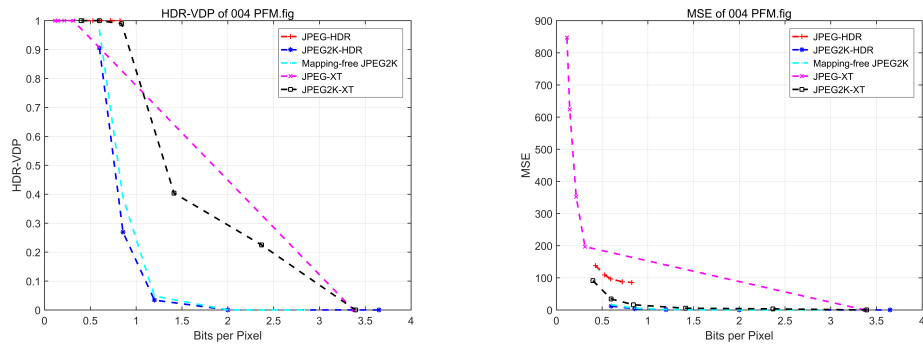


Figure A78: PFM 004.

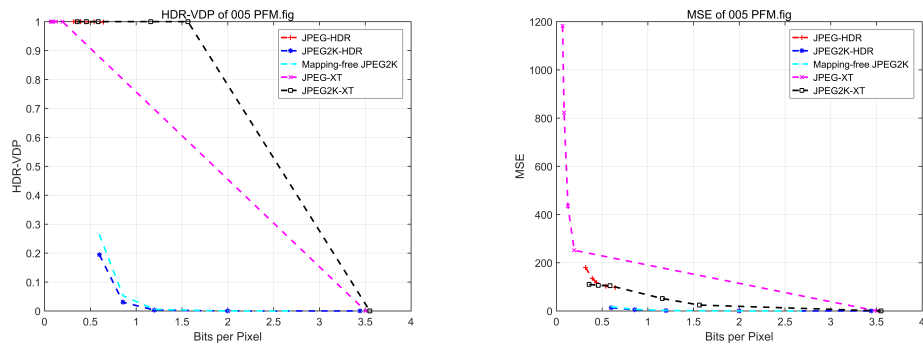


Figure A79: PFM 005.

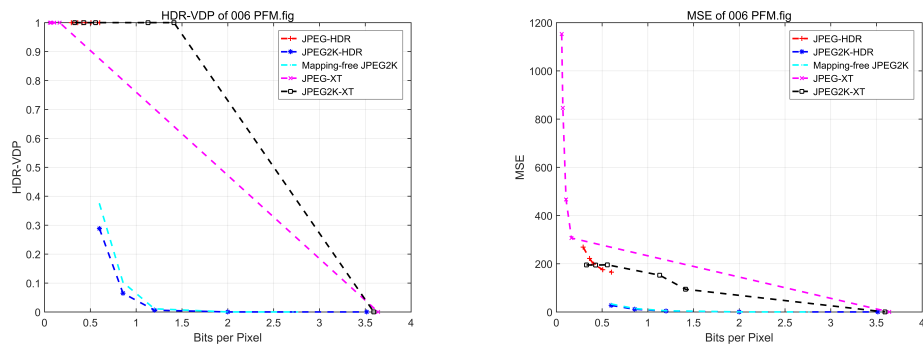


Figure A80: PFM 006.

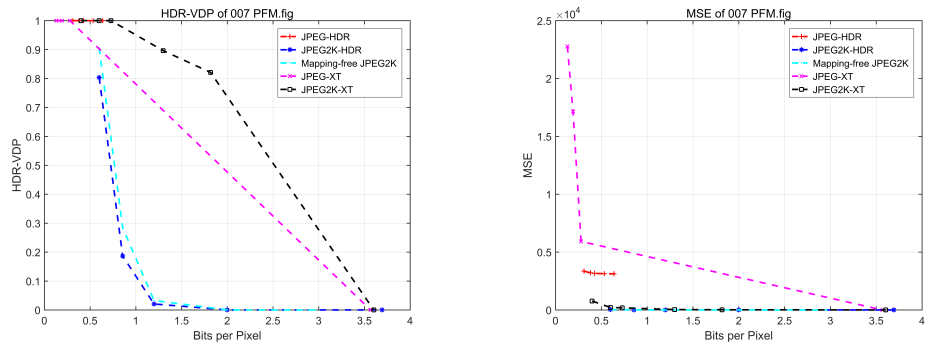


Figure A81: PFM 007.

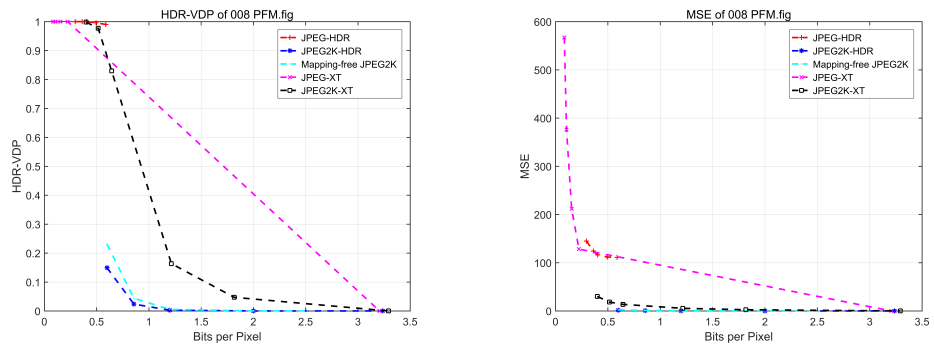


Figure A82: PFM 008.

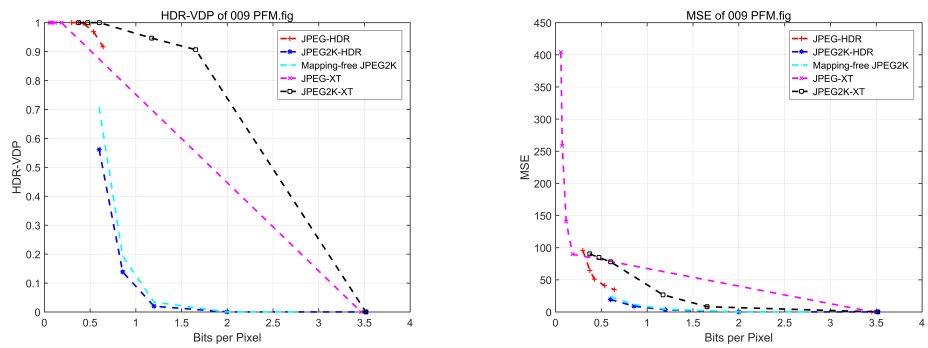


Figure A83: PFM 009.

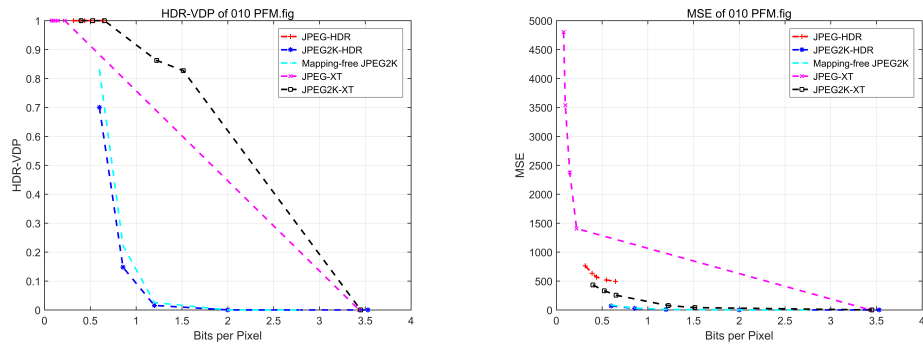


Figure A84: PFM 010.

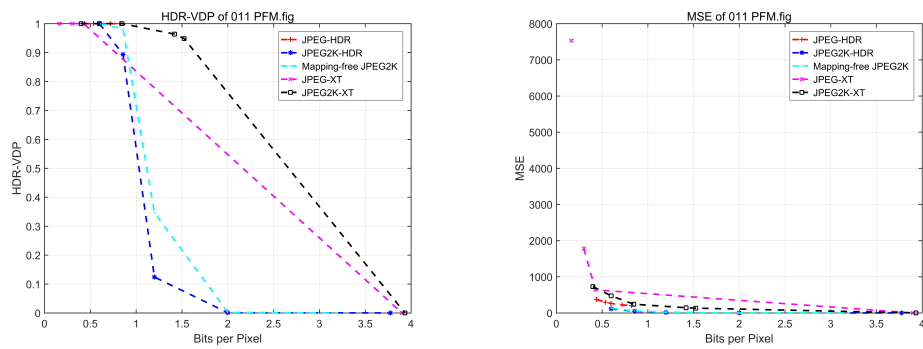


Figure A85: PFM 011.

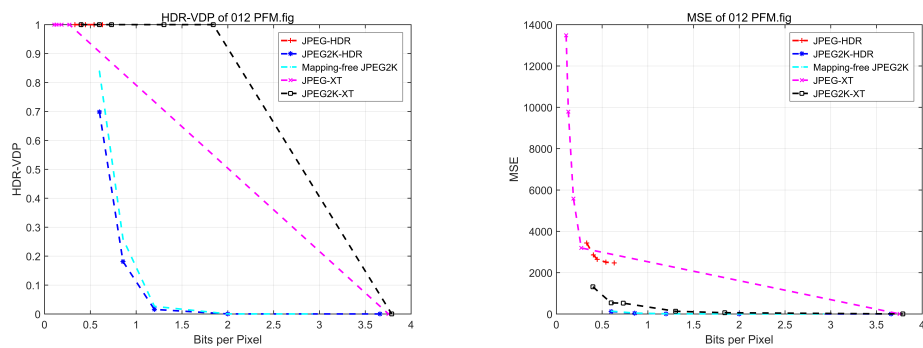


Figure A86: PFM 012.

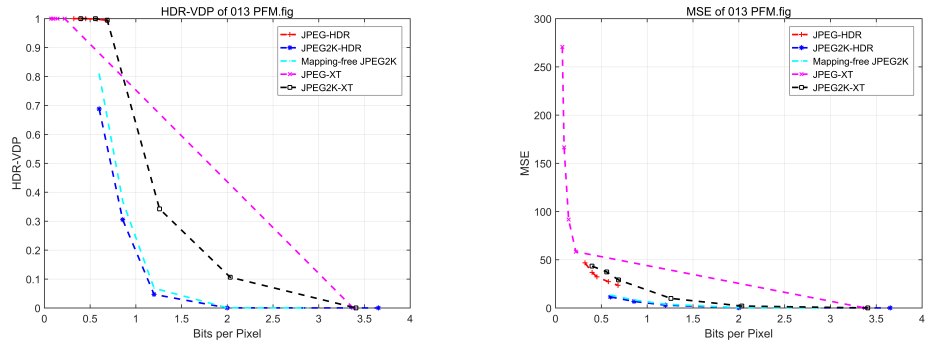


Figure A87: PFM 013.

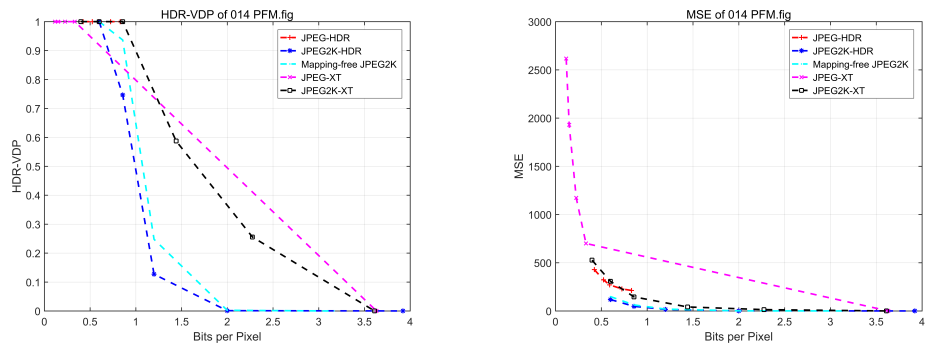


Figure A88: PFM 014.

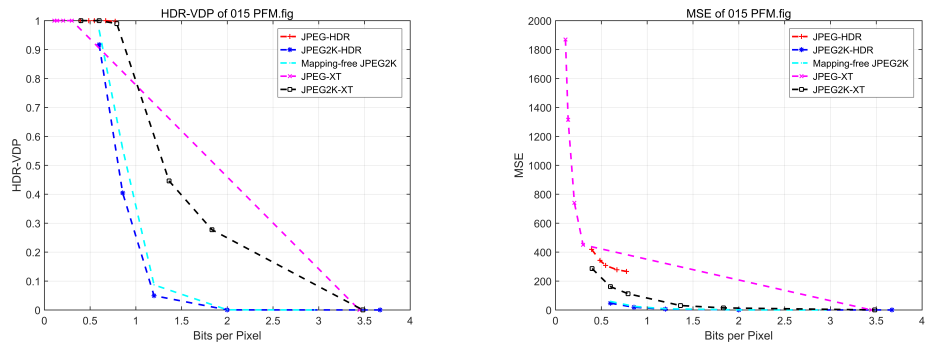


Figure A89: PFM 015.

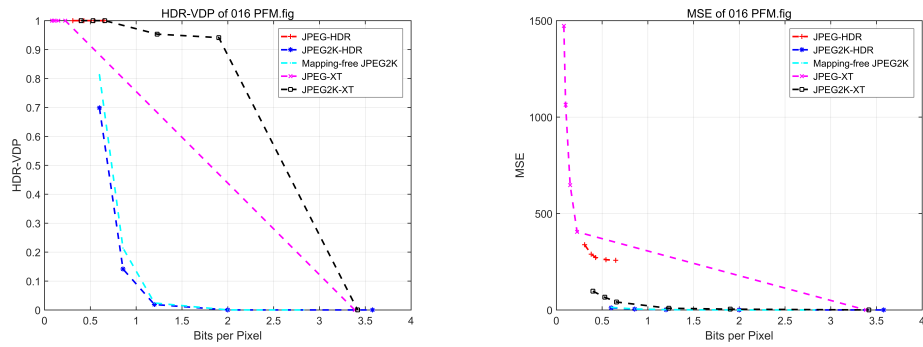


Figure A90: PFM 016.

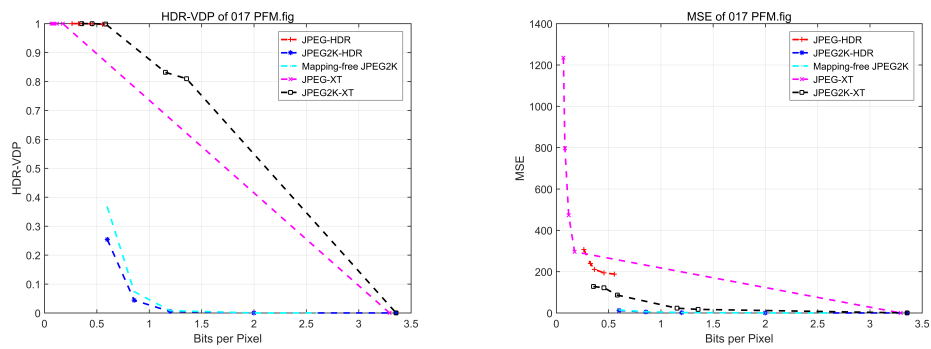


Figure A91: PFM 017.

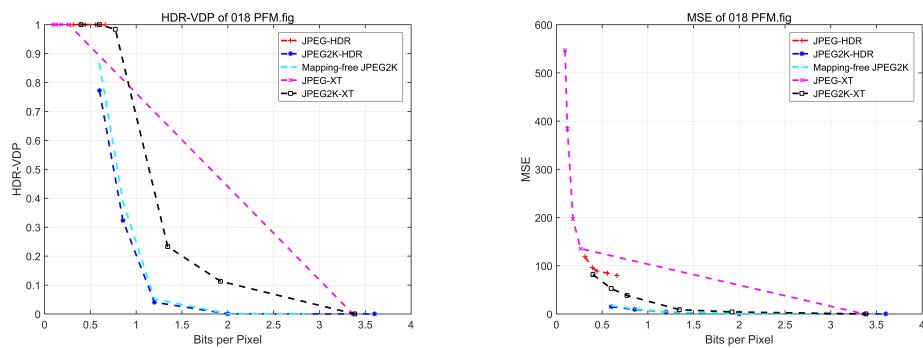


Figure A92: PFM 018.

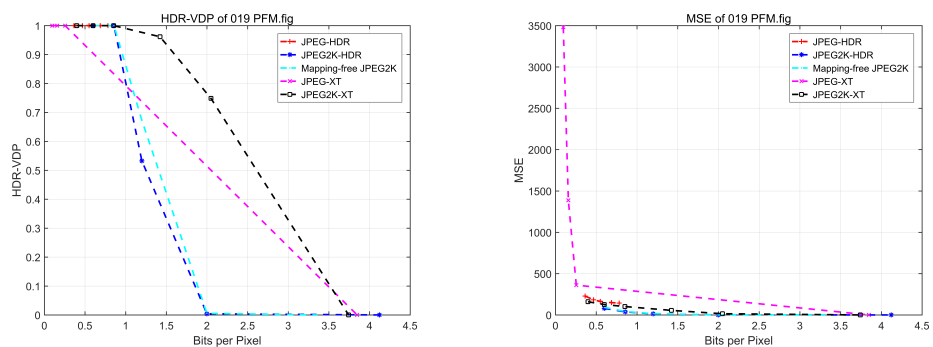


Figure A93: PFM 019.

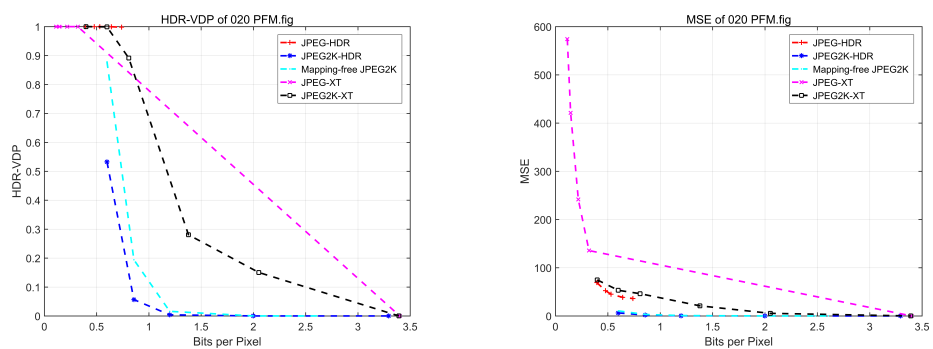


Figure A94: PFM 020.

